LONGITUDINAL DATA ANALYSIS IN DEPRESSION STUDIES: ASSESSMENT OF INTERMEDIATE-OUTCOME-DEPENDENT DYNAMIC INTERVENTIONS

by

Yen-Chih Hsu

M.S., National Yang-Ming University, Taiwan, 2003

Submitted to the Graduate Faculty of

the Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Yen-Chih Hsu

It was defended on

June 30, 2011

and approved by

Dissertation Advisor:

Abdus S. Wahed, Ph.D.
Associate Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Members:

Stewart Anderson, Ph.D.
Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Stephen R. Wisniewski, Ph.D.
Professor
Department of Epidemiology
Graduate School of Public Health
University of Pittsburgh

Joyce Chung-Chou Ho Chang, Ph.D.
Associate Professor
Department of Medicine
School of Medicine
University of Pittsburgh

LONGITUDINAL DATA ANALYSIS IN DEPRESSION STUDIES: ASSESSMENT OF INTERMEDIATE-OUTCOME-DEPENDENT DYNAMIC INTERVENTIONS

Yen-Chih Hsu, PhD

University of Pittsburgh, 2011

Longitudinal studies in the treatment of mental diseases, such as chronic forms of major depressive disorders, frequently use sequential randomization design to investigate treatment strategies. Outcomes in such studies often consist of repeated measurements of scores, such as the 24-item Hamilton Rating Scale for Depression, throughout the duration of the therapy. The goal is to compare different sequences of treatments to find the most beneficial one for each patient. Note that since treatments are applied sequentially, the eligibility of receiving one treatment assignment depends on previous treatments and outcomes. Two issues that make the analysis of data from such sequential designs different from standard longitudinal data are: (1) the randomization in the subsequent stages for patients who fail to respond in the previous stage; and (2) the drop-out of patients, for which the assumption of missing completely at random is usually not realistic. In this dissertation, we show how the inverse-probability-weighted generalized estimating equations (IPWGEE) method can be used to draw inference for treatment regimes from two-stage studies. Specifically, we show how to construct weights and use them in the IPWGEE to derive consistent estimators for the effects of treatment regimes, and compare them. Large-sample properties of the proposed estimators are derived analytically, and examined through simulations. We demonstrate our methods by applying them to a depression dataset.

Public Health Significance: Mental illness is becoming a major public health challenge. Strategies of multiple treatments have been introduced by many investigators to serve as

an alternative to single strategy in treating patients with chronic depressive disorders. As the complexity of study design increases, developing sophisticated statistical method is necessary in order to provide valid inference. This dissertation demonstrates the importance of statistical aspects to estimate the effects of depression treatment regimes from two-stage longitudinal studies.

Keywords: Counterfactual outcomes, Depression treatment regimes, Generalized estimating equations, Inverse-probability-weighting, Missing data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I owe my deepest gratitude to my advisor Dr. Abdus S. Wahed for his support and guidance. I am very fortunate to have him as my dissertation advisor and my Graduate Student Researcher (GSR) supervisor. He has made available his support in a number of ways. He is very patient and willing to help students in understanding statistical methodologies, explaining scientific results, preparing professional presentations, and building my career path.

It is a pleasure to thank my committee members Dr. Stewart Anderson, Dr. Stephen R. Wisniewski, and Dr. Joyce Chung-Chou Ho Chang for their valuable time and comments. Thanks to Dr. Anderson for sharing his experiences and caring my dissertation. This dissertation would not have been accomplished without their assistance. I especially thank Dr. Wisniewski for generously providing data from the REVAMP study that motivated my research.

I would like to thank Dr. Chang in many ways. She is not only a committee member of my dissertation but also my supervisor when I served as a teaching fellow in the Institute for Clinical Research Education at the University of Pittsburgh for three summers. It is an honor for me to be a part of her teaching team. Because of this experience, I have learned how to talk comfortably in front of the audience. That was not an easy job, but I enjoyed it very much.

I am grateful to Dr. Steven H. Belle for hiring me as a GSR in the Longitudinal Assessment of Bariatric Surgery (LABS) Consortium at the Epidemiology Data Center, University of Pittsburgh. I have learned so many analytical and communication skills from him. I also express my appreciation to the personnel in the LABS study including, but not limited to, Dr. Steven H. Belle, Dr. Abdus S. Wahed, Dr. Wendy C. King, Dr. Faith Selzer, Rocco B.

Mercurio Jr, Deborah E. Martin, Kimberly C. Beringer, Michelle A. Fouse, Laurie K. Iacono, John J. Gianakas, Abidemi K. Adeniji, Jia-Yuh Chen, and Nick Christian. It is a pleasure to work with them.

I would like to show my gratitude to all the faculty in the Department of Biostatistics at the University of Pittsburgh. They have made a great impact on my professional growth. I am heartily thankful to two past departmental chairs Dr. Howard E. Rockette and Dr. Gary M. Marsh as well as the present chair Dr. Sally C. Morton who sought funding for me throughout my doctoral education. Without their efforts, it would have been impossible to accomplish my goal comfortably. I would also like to thank Dr. Ada O. Youk for reviewing my manuscript and providing suggestions and comments.

I am indebted to many of my colleagues at the University of Pittsburgh for their support and encouragement, especially to Abidemi K. Adeniji. Abi is the most dedicated and earnest student I have even seen. His attitude has an influence on me. I feel very lucky to have a friend like Abi.

I would like to give my utmost appreciation to my family in Taiwan. My parents have given me the best education they could have possibly afforded from kindergarten to the doctorate. I owe them everything and I believe that the completion of this dissertation is the only way to show them how much I love them and appreciate their support.

No word in any language will be enough to express my debt to my fiancée Shannon Hsiang-Yu Chen for her contribution to and understanding of this achievement. We had shared happiness, anger, grief and joy together for the past four years and will continue sharing in the future. To me, she is the best gift in my life.

Finally, I sincerely appreciate all who have inspired and encouraged me to follow my dream. Today, this dream has come true. The completion of the doctorate is not the end but the beginning: new challenges are ahead of me and waiting for me. With all your inspiration and encouragement, I have all the strength to face these challenges and conquer them. I wish you all the best and hope to see you in the future.

July 18, 2011

Pittsburgh

1.0 INTRODUCTION

In this chapter, some important concepts that will be repeatedly used in this dissertation will be reviewed. The topics that will be described briefly here are:

1.1 Dynamic treatment regimes,

1.2 Sequential randomized design,

1.3 Counterfactual framework,

1.4 Generalized estimating equations (GEE),

1.5 Inverse-probability-weighting, and

1.6 REVAMP study.

1.1 DYNAMIC TREATMENT REGIMES

A dynamic treatment regime (DTR), or an adaptive treatment strategy (ATS) is a treatment rule customized for each individual based on the knowledge of time-dependent treatment, covariates and the outcome history [Lavori and Dawson, 2000]. The DTR has become common in practice for treating patients with complex diseases such as cancer [Thall et al., 2002], substance abuse [Murphy et al., 2001], and depression [Lavori et al., 2001, TenHave et al., 2003]. Under the DTR, clinicians choose an initial treatment based on patient's current and historical states, make decision for the subsequent treatment based on patient's intermediate outcome and history, and repeat the process until achieving the goal of the treatment [Lavori and Dawson, 2000]. Let $t_0, t_1, \cdots, t_K$ be times at which treatment decisions are made. Let

the data of covariates and treatments on an individual consist of L_0, A_0, L_1, A_1, L_2, A_2, $\cdots$, L_K, A_K, where

- L_0: baseline covariates measured prior to time t_0,
- A_0: treatment assignment at t_0 from a set of potential treatments $\mathcal{A}_0$,
- L_1: information including intermediate response and other covariates collected after t_0 but before t_1,
- A_1: treatment assignment at t_1 from a set of potential treatments $\mathcal{A}_1$,
- So forth for L_j and A_j, j=2,$\cdots$,K.

Also define $\bar{L}_j = (L_0, \cdots, L_j)$ and $\bar{A}_j = (A_0, \cdots, A_j)$ to be cumulative history of covariates and treatments. A DTR is a function of covariate history and prior treatments, i.e. $g : (\mathcal{T}, \bar{\mathcal{L}}_K) \to \bar{\mathcal{A}}_K$ such that for every $j = 0, \cdots, K$ and for every $\bar{l}_j \in \bar{\mathcal{L}}_j$ ($\bar{l}_j$ is a realization of $\bar{L}_j$), $g(t_j, \bar{l}_j) = a_j \in \mathcal{A}_j$. For example, in a two-stage design, the data of covariates and treatments on an individual subject consist of L_0, A_0, L_1, A_1, where

- $L_0 = \{\text{age}, \text{sex}\}$, $\mathcal{L}_0 = \{\mathbb{R}, \{0, 1\}\}$
- $A_0 \in \mathcal{A}_0 = \{a_{01}, a_{02}\}$,
- $L_1 = \{\text{Response, R}\}$, $\mathcal{L}_1 = \{0, 1\}$,
- $A_1 \in \mathcal{A}_1 = \{a_{11}, a_{12}, a'_{11}, a'_{12}\}$.

Let us consider the DTR "treat with a_{01}; if respond, treat with a_{11}; if not, treat with a'_{11}." This can be expressed in functional form as

$$g(t_0, l_0) = a_{01} \text{ (regardless of covariate values, patients receive } a_{01} \text{ at baseline)}$$

$$g(t_1, l_0, l_1) = \begin{cases} a_{11} \text{ if } l_1 = 1 \text{ (patients who respond to } a_{01} \text{ receive } a_{11} \text{ at } t_1) \\ a'_{11} \text{ if } l_1 = 0 \text{ (patients who do not respond to } a_{01} \text{ receive } a'_{11} \text{ at } t_1). \end{cases}$$

1.2 SEQUENTIAL RANDOMIZED DESIGN

A naive way to compare the effect of the treatment regimes is to randomize patients to all the possible regimes. However, the number of the decision points, the number of treatments, or patients' characteristics can cause the curse of dimensionality, i.e. the number of regimes

increases rapidly and makes it impossible to randomize patients to all the regimes. Even if we can list all possible treatment regimes, only a few of them would be viable. Thus, it is not realistic to randomize patients to all possible regimes. An alternative way is to use the sequential randomized design [Murphy, 2005] or the "play-the-winner-and-drop-the-loser" algorithm [Thall et al., 2000] to patients who meet the predetermined criteria. Compared to the naive design, the sequential randomized design with proper statistical analysis will provide more efficient and higher power of the estimator for the effect of the regimes [Ko, 2010].

1.3 COUNTERFACTUAL FRAMEWORK

In this dissertation, we use the framework of counterfactuals [Splawa-Neyman, 1990, Rubin, 1974, 1978, Robins, 1986, 1987] to quantify the effect of treatment regimes. Let us consider a non-time-varying treatment first. Let A denote the treatment, which has two values, 1 and 0. For subject i, let $Y_i(a)$ be the potential value of the outcome measurement had the subject been treated with $A = a$, $a \in (0, 1)$. In practice, a subject can receive only one of the two treatments, 0 or 1. Thus, one of these two outcomes will be observed in practice. Therefore, $\{Y_i(1), Y_i(0)\}$ are known to be counterfactuals. If each individual could receive both treatments, one could have estimated the average causal effect of the treatment by the marginal mean of the outcome

$$\hat{E}\left\{Y_i(1) - Y_i(0)\right\} = \frac{\sum_{i=1}^{N}\left\{Y_i(1) - Y_i(0)\right\}}{N} \tag{1.1}$$

Unfortunately, as stated before, these counterfactuals or potential outcomes cannot be observed for the same subject; i.e., $Y_i(1)$ and $Y_i(0)$ both cannot be observed for subject i and the observed data is (Y_i, A_i). This is the fundamental problem of causal inference [Holland, 1986]. A number of assumptions [Robins, 1997] allow one to estimate the average causal effect of the treatment from observed data. We assume that the counterfactual data on subject i do not depend on the observed or counterfactual data for any other subject [Rubin, 1978, Robins, 1997]. The first assumption is the consistency assumption (CA). The CA is used to connect

counterfactuals and observed data [Robins, 1997, Robins et al., 2000, Ko et al., 2003]. Under CA, the observed outcome measurements can be defined as $Y_i = \sum_{a=0}^{1}[I(A_i = a) \times Y_i(a)]$. The second assumption is the sequential randomization assumption (SRA) which states that the probability of receiving treatment $A_i = a$ does not depend on the counterfactual outcome $Y_i(a)$ given the predictor L_i, i.e. $Y_i(a) \perp\!\!\!\perp A_i \mid L_i$. The SRA is guaranteed in conditionally randomized studies. Note that in observational studies, since the predictor L_i is unknown, the investigators can only collect as many predictors as possible in order to approximately satisfy the SRA. Under CA and SRA, $E[Y_i(1) - Y_i(0)] = E[\bar{Y}_1 - \bar{Y}_2]$, where $\bar{Y}_a = \sum_{i=1}^{N}[I(A_i = a)Y_i] / \sum_{i=1}^{N} I(A_i = a)$ and $a \in (0, 1)$. Thus, when CA and SRA are satisfied, $E[Y_i(1) - Y_i(0)]$ can be unbiasedly estimated by $\bar{Y}_1 - \bar{Y}_2$. The last assumption is the positivity assumption which states that given the predictor L_i, the probability of being assigned to each possible treatment is greater than zero. In most of the clinical studies, the positivity is automatically true, because investigators will assign subjects into all treatments of interest.

1.4 GENERALIZED ESTIMATING EQUATIONS (GEE)

The generalized estimating equations (GEE) approach is one of the well-known statistical methods to estimate the marginal mean from longitudinal data [Liang and Zeger, 1986, Zeger and Liang, 1986]. Let $\mathbf{Y}_i = [Y_{i1}, \cdots, Y_{in_i}]^T$ be a $n_i \times 1$ vector of the outcome measurement for subject i and $E(\mathbf{Y}_i|\mathbf{X}_i) = \mathbf{X}_i\boldsymbol{\beta}$ be the marginal mean, where $\mathbf{X}_i = [\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in_i}]^T$ is a $n_i \times p$ matrix of covariates and $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters, the generalized estimating equations is given by

$$\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}_i^{-1} \{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}\} = \mathbf{0},$$

where $\mathbf{V}_i = \phi \mathbf{A}_i^{\frac{1}{2}} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{\frac{1}{2}}$; $\mathbf{R}_i(\boldsymbol{\alpha})$ is an $n_i \times n_i$ "working" correlation matrix specified by a $s \times 1$ vector $\boldsymbol{\alpha}$; $\mathbf{A}_i$ is an $n_i \times n_i$ diagonal matrix with $v_{im}(\mathbf{X}_i; \boldsymbol{\beta})$ as the m-th element, where $v_{im}(\mathbf{X}_i, \boldsymbol{\beta})$ is the assumed working variance function of Y_{im} and ϕ is the dispersion

parameter for $m \in \{1, \cdots, n_i\}$. The solution of the generalized estimating equations, $\hat{\boldsymbol{\beta}}$, can be obtained through the iterative Gauss-Newton algorithm:

$$\hat{\boldsymbol{\beta}}^{(r+1)} = \hat{\boldsymbol{\beta}}^{(r)} + \left(\sum_{i=1}^{n} \mathbf{X}_i^T \tilde{\mathbf{V}}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^T \tilde{\mathbf{V}}_i^{-1} \left\{ \mathbf{Y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}^{(r)} \right\},$$

where $\tilde{\mathbf{V}}_i = \mathbf{V}_i[\hat{\boldsymbol{\beta}}^{(r)}, \hat{\boldsymbol{\alpha}}\{\hat{\boldsymbol{\beta}}^{(r)}, \hat{\phi}(\hat{\boldsymbol{\beta}}^{(r)})\}]$.

1.5 INVERSE-PROBABILITY-WEIGHTING

The inverse-probability-weighted estimators were introduced in survey sampling by Horvitz and Thompson [Horvitz and Thompson, 1952]. To explain the underlying concept, let us assume that one is interested in estimating the population mean $\mu = \frac{1}{N} \sum_{i=1}^{N} y_i$. However, it is not possible to collect all the data points y_i for the entire population of N individuals. Usual research strategy is to select a sample of individuals, which each i-th selected individual is based on a known probability π_i. The Horvitz-Thompson estimator of the population mean μ is then $\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \Delta_i \pi_i^{-1} y_i$, where $\Delta_i = 1$ if subject i is selected and $\Delta_i = 0$ otherwise. Each selected subject is weighted by π_i^{-1} to represent him/herself and the other $\pi_i^{-1} - 1$ subjects who are not selected and have similar characteristics. By inversely weighting each selected subjects, one creates a pseudo-population and calculates the estimator of population mean as if one has collected data on the entire population. The concept of inverse-probability-weighting has been adapted by various authors in order to account for the missing data [Robins and Rotnitsky, 1992, Murphy et al., 2001, Hernán, 2004, Bembom and van der Laan, 2008]. In most clinical research, because of treatment assignment or drop-out, not all subjects will follow a specific regime. One can treat the lack of information as if the data are missing and use the inverse-probability-weighting method to estimate specific parameters of interest.

1.6 REVAMP STUDY

Between 2002 and 2006, the "Research Evaluating the Value of Augmenting Medication with Psychotherapy (REVAMP)" study [Trivedi et al., 2008] enrolled a total of 808 patients to determine the role of adjunctive psychotherapy in chronically depressed patients who had less than complete response to an initial medication. The study consisted of two 12-week stages. In the first stage, patients were assigned one of four treatments of antidepressants by the REVAMP physicians based on the algorithm using information on pharmacotherapy treatment history. These antidepressants were Sertaline (SERT), Escitalopram (EcCIT), Bupropion (BUP-SR), and Venlafaxine (VLF-XR). After up to 12 weeks of treatment in the first stage, patients not meeting certain response criteria had their pharmacotherapy modified based on the pharmacotherapy algorithm and were randomly assigned to one of three treatment strategies in the second stage: Cognitive Behavioral Analysis System of Psychotherapy (CBASP) plus medication, Brief Supportive Psychotherapy (BSP) plus medication, and Medication alone (MED). For patients with full response to their antidepressant in stage one, the same antidepressant was given continuously in stage two. In both stages, a patient's 24-item Hamilton Rating Scale for Depression (HRSD) [Hamilton, 1960], was collected at each visit.

The three specific aims of the REVAMP study were: (1) to compare the efficacy of adding psychotherapy to a medication change versus changing medication alone (MED) in chronic depressives with partial response or nonresponse to an initial trial of antidepressant medication, (2) to test efficacy of the CBASP as an augmentation strategy by comparing it to BSP, and (3) to test a hypothesized mechanism of therapeutic action of CBASP by examining whether patients receiving CBASP exhibit significantly greater improvements in social problem solving than patients receiving BSP or MED. The results from several studies related to the REVAMP study had suggested the need for multi-stage of sequential treatments in order to achieve a response [Trivedi et al., 2008, Kocsis et al., 2008, 2009, Klein et al., 2009].

1.7 MOTIVATION AND OBJECTIVES

Sequential randomization design is common in clinical studies. Conventional analysis strategy for data from multi-stage design is to analyze data in separate stages. For example, in a two-stage design, an initial treatment is given in the first stage and a maintenance (or alternate) treatment is given in the second stage depending on the patient's response status after the initial treatment. The estimation and comparison between two maintenance (or alternate) treatments is usually done by conditioning on response to the initial treatment. This can answer, for example, the question of which alternative treatment is the best for those who did not respond to the initial treatment. But it does not allow us to answer questions, such as "which treatment regime is the best for given individual?" In this dissertation, we propose methods for estimating the effects of dynamic treatment regimes from longitudinal studies by using the generalized estimating equations and the inverse-probability-weighting methods.

One of the challenges in analyzing data from longitudinal studies is the presence of missing data. Most statistical methods for longitudinal data are valid as long as the data are missing completely at random (MCAR). However, the MCAR assumption may not be realistic in clinical studies. In contrast, the assumption of missing at random (MAR) is more persuasive and the unbiased estimates can be obtained through advanced statistical methods, e.g. inverse-probability-weighting method.

The objective of this dissertation is two-fold: (1) propose statistical methods which properly accounts for missing data issues for estimating and comparing the effects of dynamic treatment regimes in longitudinal studies and (2) improve the efficiency of the proposed estimators by incorporating partially observed information from individuals who have missing data due to randomization or drop-out.

2.0 WEIGHTED GEE FOR RESPONSE-ADAPTIVE TREATMENT REGIMES IN TWO-STAGE LONGITUDINAL STUDIES

2.1 INTRODUCTION

In the past decade, a number of studies have shown the efficacy of pharmacotherapies and psychotherapies in treatment of chronic forms of major depressive disorders (cMDD). However, nearly 50% of patients with cMDD fail to respond to the first line pharmacotherapies or psychotherapies [Kocsis et al., 2009]. Hence, for those who do not respond adequately to the first line therapy, combining pharmacotherapy and psychotherapy to replace monotherapy is becoming more and more frequent in clinical practice [Cuijpers et al., 2009]. Combination treatment, in general, is more expensive than monotherapy. A practical treatment strategy could be initially giving cMDD patients a common medication, which is usually cheaper, such as an antidepressant, and adding psychotherapy, which is more expensive, if patients have poor or partial response to the initial medication. Often, multiple stages of treatments may be necessary to achieve a response. A patient moves to the next stage of therapy in two circumstances: (1) if a patient achieves response, the patient could continue the same therapy or his/her therapy could be modified to maintain the response; and (2) if a patient fails to respond in the previous stage, the patient would be given some alternative therapies to achieve a response.

In an attempt to investigate the effect of sequence of pharmacotherapy and psychotherapy in the treatment of patients with chronic depression, the Research Evaluating the Value of Augmenting Medication with Psychotherapy (REVAMP) study [Trivedi et al., 2008] enrolled a total of 808 patients to determine the role of adjunctive psychotherapy in chronically depressed patients who had less than complete response to an initial medication. The

study consisted of two 12-week stages. In the first stage, patients were assigned one of four treatments of antidepressants by the REVAMP physicians based on the algorithm using information on pharmacotherapy treatment history. These antidepressants were Sertaline (SERT), Escitalopram (EcCIT), Bupropion (BUP-SR), and Venlafaxine (VLF-XR). After up to 12 weeks of treatment in the first stage, patients not meeting certain response criteria had their pharmacotherapy modified based on the pharmacotherapy algorithm and were randomly assigned to one of three treatment strategies in the second stage: Cognitive Behavioral Analysis System of Psychotherapy (CBASP) plus medication, Brief Supportive Psychotherapy (BSP) plus medication, and Medication alone (MED). For patients with full response to their antidepressant in the first stage, the same antidepressant was given continuously in the second stage. In both stages, a patient's 24-item Hamilton Rating Scale for Depression (HRSD) score [Hamilton, 1960], was collected at each visit.

The purpose of the REVAMP study was to determine optimal adjunctive psychotherapies with which chronically depressed patients would benefit most. Since the adjunctive psychotherapies would be offered only if a chronically depressed patient fails to respond to the pharmacotherapy, possible choices of treatment regimes would be to continue the same pharmacotherapy if a patient responds to the therapy and to choose a psychotherapy if he/she fails to respond to the pharmacotherapy. Since there would be different options of pharmacotherapies and psychotherapies, the practicing physicians would have to choose from many different treatment regimes. For example, in the REVAMP study, there were three possible treatment regimes: (1) treat with an antidepressant, continue the antidepressant if respond, otherwise add CBASP to the antidepressant; (2) treat with an antidepressant, continue the antidepressant if respond, otherwise add BSP to the antidepressant; and (3) treat with an antidepressant, continue the antidepressant if respond, otherwise add MED alone to the antidepressant. The aim of this study was to compare the efficacy of adding CBASP or BSP to continued treatment of antidepressants with MED alone. One naive way to analyze data from such designs would be to compare the psychotherapies conditional on the fact that the patients did not respond in stage I. However, this would be a conditional analysis and would not address the question of choosing an overall treatment regime to be best among all possible regimes.

The generalized estimating equations (GEE) approach is one of the well-known statistical methods to estimate the marginal treatment effect from longitudinal data [Liang and Zeger, 1986, Zeger and Liang, 1986]. In a study with two stages of therapies, such as the REVAMP study, a patient can belong to several treatment regimes. For example, patients who had responded to the initial treatment would continue their treatment in the second stage in the REVAMP study. These patients were treated consistently with all three regimes. Therefore, the standard GEE may not be directly applicable. Additionally, patients who become eligible for the second stage treatments receive treatment by randomization and hence, the inverse-probability-weighted GEE (IPWGEE) [Robins et al., 1995] can be used to account for the randomization. Moreover, patients may drop out throughout the study. Drop-outs may be due to lack of efficacy to the pharmacotherapy or psychotherapy. Ignoring these patients in estimation might lead to biased estimates of effects of treatment regimes [Diggle et al., 2002]. In this study, we use the IPWGEE to account for drop-outs. We derive consistent and asymptotically normal estimators of regime effects, and provide the Wald test for comparing different regimes.

This chapter is organized as follows. We introduce notation, model, and assumptions in Section 2.2. In Section 2.3, we show how to draw inference from complete data (no drop-out.) Section 2.4 modifies the estimators from Section 2.3 to account for the drop-outs. The Wald test statistic for comparing treatment regimes is given in Section 2.5. In Section 2.6, we evaluate the large-sample properties of the proposed methods through simulations. In Section 2.7, we demonstrate these methods through an application to the REVAMP dataset. We wrap up with a discussion in Section 2.8. The figures and tables are listed in Section 2.9.

2.2 DATA, MODEL, AND ASSUMPTIONS

We consider a design that is more general than the REVAMP study (Figure 2.1). We assume that each patient i ($i = 1, \ldots, n$) randomly receives a first line treatment A_j in the first stage, where $j \in \{1, \ldots, J\}$. Patients who respond to A_j are randomized to a maintenance treatment B_k, where $k \in \{1, \ldots, K\}$, and patients who do not respond to A_j

are randomized to an alternative treatment B'_l, where $l \in \{1, \ldots, L\}$. The objective of this study is to estimate and compare the effects of various treatment regimes arising from a combination of the initial treatment, intermediate response, and the second stage treatment. A treatment regime, $A_j B_k B'_l$, is defined as "treat with A_j followed by B_k if respond, by B'_l if otherwise." Patients are followed over time, and for patient i, a continuous outcome Y_{im} (e.g. the 24-item Hamilton Rating Scale for Depression) is measured at time t_{im}, where $m \in \{1, \ldots, M_i\}$.

In the presence of randomization and drop-out, it is often useful to apply the idea of counterfactuals to the data analysis [Holland, 1986]. For patient i ($i = 1, \ldots, n$), define $R_i(A_j)$ to be the response status if the patient receives the first line treatment A_j; let $T_{im_{1i}}$ be the time when patient i is declared a responder or non-responder to the first line treatment A_j, at which point randomization to the second set of treatment occurs, $m_{1i} \in \{1, \ldots, M_i\}$. Whether observed or not, we define the following outcomes: $\mathbf{Y}_i(A_j)$, a $m_{1i} \times 1$ vector of repeated measures of outcome at time points in the first stage if patient i receives treatment A_j in the first stage; $\mathbf{Y}_i(A_j B_k)$, a $(M_i - m_{1i}) \times 1$ vector of repeated measures of outcome at each time point in the second stage if patient i receives treatment A_j in the first stage and B_k in the second stage after responding to A_j; $\mathbf{Y}_i(A_j B'_l)$, a $(M_i - m_{1i}) \times 1$ vector of repeated measures of outcome at each time point in the second stage if patient i receives treatment A_j in the first stage and B'_l in the second stage after failing to respond to A_j. For simplicity, let us assume $J = K = L = 2$. Thus, for one initial treatment A_1, patient i could be associated with the following random variables:

$$\{[R_i(A_1), T_{im_{1i}}], \mathbf{Y}_i(A_1), \mathbf{Y}_i(A_1 B_1), \mathbf{Y}_i(A_1 B_2), \mathbf{Y}_i(A_1 B'_1), \mathbf{Y}_i(A_1 B'_2)\}. \tag{2.1}$$

In terms of these counterfactual variables, let $\mathbf{Y}_i(A_1 B_k B'_l)$ be the $M_i \times 1$ vector outcome of patient i under treatment regime $A_1 B_k B'_l$, for $k, l \in \{1, 2\}$. The m-th element of $\mathbf{Y}_i(A_1 B_k B'_l)$ represents the outcome of patient i at time t_{im} under regime $A_1 B_k B'_l$, and it can be expressed as:

$$\begin{aligned} Y_{im}(A_1 B_k B'_l) &= I(t_{im} \leq T_{im_{1i}}) Y_{im}(A_1) \\ &\quad + I(t_{im} > T_{im_{1i}}) \left\{ R_i Y_{i(m-m_{1i})}(A_1 B_k) + (1 - R_i) Y_{i(m-m_{1i})}(A_1 B'_l) \right\}. \end{aligned} \tag{2.2}$$

In a similar fashion, we can define the outcome, $Y_{im}(A_2B_kB_l')$, for patient i at time t_{im} given $A_2B_kB_l'$. Therefore, without loss of generality, we will consider one treatment A_1 in the first stage. Our interest is to estimate the effect of treatment regime $Y_{im}(A_1B_kB_l')$ over time, which is formulated as the coefficient $\beta_{1,1kl}$ of time in the marginal mean model:

$$E\left[Y_{im}(A_1B_kB_l') \mid \mathbf{x}_{im}\right] = \mathbf{x}_{im}^T\boldsymbol{\beta}_{1kl}, \tag{2.3}$$

where $\mathbf{x}_{im}^T = [1, t_{im}, w_{1i}, \ldots, w_{pi}]$ and $w_{1i}, \ldots, w_{pi}$ are p baseline covariates for patient i. In other words, we would like to estimate the parameters $\boldsymbol{\beta}_{1kl} = [\beta_{0,1kl}, \ldots, \beta_{p+1,1kl}]^T$. If $Y_{im}(A_1B_kB_l')$ were observed for each patient in the sample, the generalized estimating equations could have been used to estimate these coefficients. However, in reality, we cannot observe the outcome $Y_{im}(A_1B_kB_l')$ for all patients. For example, if a patient receives A_1, responds to A_1, and then receives $B_{k'}$ where $k' \neq k$, we do not observe $Y_{im}(A_1B_kB_l')$ for that patient. The complete observed data are characterized as the set of i.i.d. random list:

$$\left\{[R_i, T_{im_{1i}}], R_iZ_{ki}, (1-R_i)Z_{li}', \mathbf{W}_i, \mathbf{Y}_i\right\}, k, l = 1, 2, i = 1, \ldots, n,$$

where $R_i = 1$, if the patient is a responder, and $R_i = 0$, if otherwise; $\mathbf{W}_i = [W_{1i}, \ldots, W_{pi}]^T$ is a $p \times 1$ vector of p baseline covariates; Z_{ki} and Z_{li}' are the assignment indicators for treatment B_k and B_l', respectively, for $k, l \in \{1, 2\}$; $Z_{1i} = 1(0)$ if patient i is randomized to $B_1(B_2)$; $Z_{1i}' = 1(0)$ if patient i is randomized to $B_1'(B_2')$; Z_{2i} and Z_{2i}' satisfy $Z_{1i} + Z_{2i} = 1$ and $Z_{1i}' + Z_{2i}' = 1$; $\mathbf{Y}_i$ is a $M_i \times 1$ vector of repeated observed outcome for patient i.

In order to draw inference on $\mathbf{Y}_i(A_1B_kB_l')$ from observed data, the consistency assumption (CA) is required to connect observed data and counterfactuals [Rubin, 1974, Robins et al., 2000]. The CA implies that the observed outcome is equal to the counterfactual outcome under treatment assignment consistent with the counterfactual. In other words, for $m \in \{1, 2, \ldots, M_i\}$,

$$\begin{aligned} Y_{im} &= I(t_{im} \leq T_{im_{1i}})Y_{im}(A_1) \\ &\quad + I(t_{im} > T_{im_{1i}})\left\{R_i\sum_{k=1}^{2} Z_{ki}Y_{i(m-m_{1i})}(A_1B_k) + (1-R_i)\sum_{l=1}^{2} Z_{li}'Y_{i(m-m_{1i})}(A_1B_l')\right\}. \end{aligned} \tag{2.4}$$

Another frequently made assumption is the Sequential Randomization Assumption (SRA) which states that the probabilities of receiving treatment B_k and B'_l do not depend on counterfactuals given the history of information collected prior to the randomization [Rubin, 1974, Robins, 1986]:

$$P\left\{Z_{ki}=1 \mid [R_i=1, T_{im_{1i}}], \mathbf{W}_i, \mathbf{Y}_i(A_1B_kB'_l)\right\} = P\left\{Z_{ki}=1 \mid [R_i=1, T_{im_{1i}}], \mathbf{W}_i\right\};$$
$$P\left\{Z'_{li}=1 \mid [R_i=0, T_{im_{1i}}], \mathbf{W}_i, \mathbf{Y}_i(A_1B_kB'_l)\right\} = P\left\{Z'_{li}=1 \mid [R_i=0, T_{im_{1i}}], \mathbf{W}_i\right\};$$
$$\text{for } k,l \in \{1,2\}. \tag{2.5}$$

In the REVAMP study, the probabilities of treatment assignment in the second stage were constant. Therefore, we define $P\{Z_{ki}=1 \mid [R_i=1, T_{im_{1i}}], \mathbf{W}_i\} = \eta_k$ and $P\{Z'_{li}=1 \mid [R_i=0, T_{im_{1i}}], \mathbf{W}_i\} = \zeta_l$.

2.3 INFERENCE FROM COMPLETE DATA: NO DROP-OUT

If everyone in our sample had followed the same treatment regime $A_1B_kB'_l$, where $k,l \in \{1,2\}$, we could have used the GEE method [Liang and Zeger, 1986, Zeger and Liang, 1986] to estimate the regime effect, i.e. a $(p+2)\times 1$ vector of parameter $\boldsymbol{\beta}_{1kl} = [\beta_{0,1kl}, \beta_{1,1kl}, \ldots, \beta_{p+1,1kl}]^T$. The generalized estimating equations in this case would be given by

$$\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \left\{\mathbf{Y}_i(A_1B_kB'_l) - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right\} = \mathbf{0}, \tag{2.6}$$

where $\mathbf{X}_i^T = [\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iM_i}]$, and $\mathbf{x}_{im}^T = [1, t_{im}, \mathbf{W}_i^T]$, where $m \in \{1,2,\ldots,M_i\}$; $\mathbf{R}(\boldsymbol{\alpha})$ is an $M_i \times M_i$ "working" correlation matrix specified by $\boldsymbol{\alpha}$; $\mathbf{A}_{i,1kl}$ is an $M_i \times M_i$ diagonal matrix with $v_{im}(\mathbf{X}_i;\boldsymbol{\beta}_{1kl})$ as the m-th element, where $v_{im}(\mathbf{X}_i,\boldsymbol{\beta}_{1kl})$ is the assumed working variance function of $Y_{im}(A_1B_kB'_l)$ and ϕ is the dispersion parameter; $\mathbf{V}_{i,1kl} = \phi\mathbf{A}_{i,1kl}^{\frac{1}{2}}\mathbf{R}(\boldsymbol{\alpha})\mathbf{A}_{i,1kl}^{\frac{1}{2}}$.

However, not all patients followed the treatment regime $A_1B_kB'_l$. Some patients in the study received treatments inconsistent with $A_1B_kB'_l$, i.e. these patients randomized to receive other second stage treatment B_{3-k} or B'_{3-l}, where $k,l \in \{1,2\}$. The data from these patients can be treated as missing data while estimating $\boldsymbol{\beta}_{1kl}$. Because of randomization,

these patients with treatments inconsistent with $A_1B_kB'_l$ are similar to those treated under treatment regime $A_1B_kB'_l$. Thus, the inverse-probability-weighted method [Horvitz and Thompson, 1952, Rosenbaum and Rubin, 1983] can be used to account for the data that are missing by randomization. Patients randomized to treatment B_k are weighted by $1/\eta_k$ and patients randomized to treatment B'_l are weighted by $1/\zeta_l$. This way, a patient randomized to treatment B_k or B'_l counts for him/herself as well as for $(1/\eta_k - 1)$ or $(1/\zeta_k - 1)$ similar patients who have "missing data" with respect to treatment regime $A_1B_kB'_l$ (i.e. randomized to a second stage treatment other than B_k or B'_l.) This inverse probability of treatment weighting will be applied to create a pseudo-population from each patient who follows the policy $A_1B_kB'_l$. The weight, $\mathbf{Q}_{i,1kl}$, is thus defined to be an $M_i \times M_i$ diagonal matrix with each m-th diagonal element, $Q_{im,1kl}$, defined as

$$\left[\frac{R_i Z_{ki}}{\eta_k} + \frac{(1-R_i)Z'_{li}}{\zeta_l}\right].$$

Given the treatment regime $A_1B_kB'_l$, we use the IPWGEE method [Robins et al., 1995] with the "weight" $\mathbf{Q}_{i,1kl}$ to estimate the regime effect $\boldsymbol{\beta}_{1kl}$. The weighted estimating equation is then

$$\sum_{i=1}^{n} \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}\right) = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \mathbf{Q}_{i,1kl}\left\{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right\} = \mathbf{0}, \tag{2.7}$$

where $\mathbf{Y}_i$, $\mathbf{X}_i$, and $\mathbf{V}_{i,1kl}$ are defined as in (2.6). The solution of (2.7), $\hat{\boldsymbol{\beta}}_{1kl}$, can be obtained through the following iterative algorithm [Liang and Zeger, 1986]:

$$\hat{\boldsymbol{\beta}}_{1kl}^{(r+1)} = \hat{\boldsymbol{\beta}}_{1kl}^{(r)} + \left(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \mathbf{Q}_{i,1kl} \mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \mathbf{Q}_{i,1kl}\left\{\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_{1kl}^{(r)}\right\}. \tag{2.8}$$

Lemma 2.3.1. *Under SRA in (2.5),* $E[\mathbf{Q}_{i,1kl} \mid \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB'_l)] = \mathbf{I}_{M_i}$.

Proof. This follows since

$$\begin{aligned}
&E\left[Q_{im,1kl} \mid \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB'_l)\right] \\
&= E\left\{E\left[\frac{R_i Z_{ki}}{\eta_k} + \frac{(1-R_i)Z'_{li}}{\zeta_l} \mid R_i, \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB'_l)\right]\right\} \\
&= E\left\{\frac{R_i}{\eta_k} E\left[Z_{ki} \mid R_i, \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB'_l)\right] + \frac{1-R_i}{\zeta_l} E\left[Z'_{li} \mid R_i, \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB'_l)\right]\right\} \\
&= 1 \qquad \text{(by SRA), for } m \in \{1, \ldots, M_i\}.
\end{aligned}$$

□

Proposition 2.3.1. *Under CA in (2.4) and SRA in (2.5),* $\hat{\boldsymbol{\beta}}_{1kl}$ *is a consistent estimator of* $\boldsymbol{\beta}_{1kl}$.

Proof. First note that by CA in (2.4), $\mathbf{Q}_{i,1kl}\mathbf{Y}_i = \mathbf{Q}_{i,1kl}\mathbf{Y}_i(A_1B_kB_l')$. Also note that $\hat{\boldsymbol{\beta}}_{1kl}$ satisfies Equation (2.7), and therefore, to show that $\hat{\boldsymbol{\beta}}_{1kl}$ is consistent, it suffices to show that $E[\mathbf{U}_i(\boldsymbol{\beta}_{1kl})] = \mathbf{0}$. Now,

$$\begin{aligned}
E\left[\mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}\right)\right] &= E\left[\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}\left\{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right\}\right] \\
&= E\left[\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}\left\{\mathbf{Y}_i(A_1B_kB_l') - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right\}\right] \quad \text{(by CA in (2.4))} \\
&= E\left\{E\left[\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}\left\{\mathbf{Y}_i(A_1B_kB_l') - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right\} \mid \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB_l')\right]\right\} \\
&= E\left[\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\left\{\mathbf{Y}_i(A_1B_kB_l') - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right\}\right] \quad \text{(by Lemma 2.3.1)} \\
&= E\left\{E\left[\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\left\{\mathbf{Y}_i(A_1B_kB_l') - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right\} \mid \mathbf{X}_i\right]\right\} \\
&= E\left\{\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\left\{E\left[\mathbf{Y}_i(A_1B_kB_l') \mid \mathbf{X}_i\right] - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right\}\right\} = \mathbf{0} \quad \text{(by Equation (2.3))}.
\end{aligned}$$

□

Proposition 2.3.2. *Under CA in (2.4) and SRA in (2.5),* $\hat{\boldsymbol{\beta}}_{1kl}$ *is asymptotically normally distributed with mean* $\boldsymbol{\beta}_{1kl}$ *and variance* Σ/n*, where*

$$\left.\begin{aligned}
&\boldsymbol{\Sigma} = \mathbf{C}^{-1}(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl})\mathbf{B}(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl})\mathbf{C}^{-1}(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}), \\
&\mathbf{C}(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}) = E\left[\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{X}_i\right], \text{ and} \\
&\mathbf{B}(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}) = E\left[\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}\left\{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right\}\left\{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right\}^T\mathbf{Q}_{i,1kl}\mathbf{V}_{i,1kl}^{-1}\mathbf{X}_i\right].
\end{aligned}\right\} \quad (2.9)$$

Proof. We first note that the estimator $\hat{\boldsymbol{\beta}}_{1kl}$ satisfies $\sum_{i=1}^n \mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}\{\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_{1kl}\} = \mathbf{0}$. Expanding $\sum_{i=1}^n \mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}\{\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_{1kl}\}$ around $\boldsymbol{\beta}_{1kl}$ using Taylor's expansion, we obtain

$$\begin{aligned}
&n^{\frac{1}{2}}\left(\hat{\boldsymbol{\beta}}_{1kl} - \boldsymbol{\beta}_{1kl}\right) \\
&= \left[\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}\mathbf{X}_i\right]^{-1} n^{-\frac{1}{2}}\sum_{i=1}^n \mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}\left\{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right\} + \mathbf{o_p}(\mathbf{1}), \qquad (2.10)
\end{aligned}$$

where $\mathbf{o_p}(\mathbf{1})$ is a term that converges in probability to $\mathbf{0}$ as $n \to \infty$. Note also that

$$\left[\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}\mathbf{X}_i\right] \xrightarrow{p} E\left[\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}\mathbf{X}_i\right] = E\left[\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{X}_i\right] = \mathbf{C}\left(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}\right).$$

Therefore, from Equation (2.10), one can write

$$n^{\frac{1}{2}}\left(\hat{\boldsymbol{\beta}}_{1kl}-\boldsymbol{\beta}_{1kl}\right)=n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{\psi}_{i,1kl}+\mathbf{o_p}(\mathbf{1}), \tag{2.11}$$

where $\boldsymbol{\psi}_{i,1kl}=\mathbf{C}^{-1}(\phi,\boldsymbol{\alpha},\boldsymbol{\beta}_{1kl})\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}\{\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}_{1kl}\}$ is known as the influence function of the estimator $\hat{\boldsymbol{\beta}}_{1kl}$ (where $E(\boldsymbol{\psi}_{i,1kl})=\mathbf{0}$ and $\mathrm{E}(\boldsymbol{\psi}_{i,1kl}\boldsymbol{\psi}_{i,1kl}^T)$ is positive definite.) From (2.11), we see that $\hat{\boldsymbol{\beta}}_{1kl}$ is an asymptotically linear estimator of $\boldsymbol{\beta}_{1kl}$. Applying the central limit theorem to (2.11), we deduce that $n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_{1kl}-\boldsymbol{\beta}_{1kl})\xrightarrow{d}\mathrm{MVN}(\mathbf{0},\boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}=E[\boldsymbol{\psi}_{i,1kl}\boldsymbol{\psi}_{i,1kl}^T]$ is given in (2.9). □

The asymptotic variance-covariance matrix of $\hat{\boldsymbol{\beta}}_{1kl}$ can be estimated by the empirical estimator

$$\widehat{\mathrm{var}}\left(\hat{\boldsymbol{\beta}}_{1kl}\right)=\frac{1}{n}\widehat{E}\left[\boldsymbol{\psi}_{i,1kl}\boldsymbol{\psi}_{i,1kl}^T\right]=\frac{1}{n^2}\sum_{i=1}^{n}\hat{\boldsymbol{\psi}}_{i,1kl}\hat{\boldsymbol{\psi}}_{i,1kl}^T, \tag{2.12}$$

or by the model-based estimator

$$\widehat{\mathrm{var}}\left(\hat{\boldsymbol{\beta}}_{1kl}\right)=\frac{1}{n}\left[\mathbf{C}_n^{-1}(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl})\mathbf{B}_n(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl})\mathbf{C}_n^{-1}(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl})\right], \tag{2.13}$$

where

$$\hat{\boldsymbol{\psi}}_{i,1kl}=\mathbf{C}_n^{-1}(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl})\mathbf{X}_i^T\hat{\mathbf{V}}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}\{\mathbf{Y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}_{1kl}\},$$

$$\mathbf{C}_n(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl})=\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i^T\hat{\mathbf{V}}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}\mathbf{X}_i,$$

$$\mathbf{B}_n(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl})=\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i^T\hat{\mathbf{V}}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}\{\mathbf{Y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}_{1kl}\}\{\mathbf{Y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}_{1kl}\}^T\mathbf{Q}_{i,1kl}\hat{\mathbf{V}}_{i,1kl}^{-1}\mathbf{X}_i,$$

$$\hat{\mathbf{V}}_{i,1kl}=\hat{\phi}\hat{\mathbf{A}}_{i,1kl}^{\frac{1}{2}}\mathbf{R}(\hat{\boldsymbol{\alpha}})\hat{\mathbf{A}}_{i,1kl}^{\frac{1}{2}},$$

$$\hat{\mathbf{A}}_{i,1kl}=\mathrm{diag}(v_{im}(\mathbf{X}_i;\hat{\boldsymbol{\beta}}_{1kl}),m=1,\ldots,M_i),$$

$$\hat{\phi}=\frac{\sum_{i=1}^{n}\sum_{m=1}^{M_i}\hat{e}_{im}^2}{\sum_{i=1}^{n}\sum_{m=1}^{M_i}Q_{im,1kl}-(p+2)},\ \text{and}$$

$$\hat{e}_{im}=\sqrt{Q_{im,1kl}}(Y_{im}-\mathbf{x}_{im}^T\hat{\boldsymbol{\beta}}_{1kl}).$$

The specific estimator of $\boldsymbol{\alpha}$ depends on the working correlation structure, $\mathbf{R}(\boldsymbol{\alpha})$ [Liang and Zeger, 1986]. For example, if $\mathrm{corr}(Y_{im}(A_1B_kB_l'),Y_{i(m+s)}(A_1B_kB_l'))=\alpha^s$ for $s=0,1,2,\ldots,(M_i-m)$, i.e. $\mathbf{R}(\alpha)$ has an autoregressive correlation structure (AR(1)), then $\hat{\alpha}=\frac{\sum_{i=1}^{n}\sum_{m<M_i-1}\hat{e}_{im}\hat{e}_{i(m+1)}}{[N^*-(p+2)]\hat{\phi}}$, where $N^*=\sum_{i=1}^{n}[\sum_{m=1}^{M_i}I(Q_{im,1kl}>0)-1]$.

2.4 INFERENCE FROM INCOMPLETE DATA: PRESENCE OF DROP-OUT

In Section 2.3, we used the IPWGEE method to deal with missing data due to randomization in the second stage of the therapy. We assumed that there was no drop-out. However, in longitudinal studies, drop-out is a common phenomenon. The GEE or IPWGEE can provide valid estimates of parameters as long as drop-outs are missing completely at random, MCAR [Robins et al., 1995]. If drop-outs depend on observed data (e.g. responses from previous visits or baseline characteristics), they are not MCAR but missing at random, MAR [Little and Rubin, 2002]. Therefore, we need to adjust our estimators from Section 2.3 to account for MAR.

In the presence of drop-outs, the observed data from patient i are

$$\{\mathbf{W}_i, \Delta_i, \mathbf{H}_i\}, k, l = 1, 2, i = 1, \ldots, n,$$

where $\Delta_i = 0$ if patient i had dropped out from the study before completion, and $\Delta_i = 1$ if otherwise. When $\Delta_i = 1$, $\mathbf{H}_i = \{[R_i, T_{im_{1i}}], R_i Z_{ki}, (1 - R_i) Z'_{li}, \mathbf{Y}_i\}$. When $\Delta_i = 0$, $\mathbf{H}_i$ contains post-baseline information observed prior to drop-out. Thus, when $\Delta_i = 0$ and drop-out occurs at time $t_{i(d_i+1)}$, $\mathbf{Y}_i$ is a $d_i \times 1$ vector, where $d_i < M_i$. To account for the data that are missing due to drop-out, one can extend the idea of inverse-probability-weighting method described in Section 2.3. We define the probability of a patient having complete data as $\pi_i = P\{\Delta_i = 1\}$. Had this probability been known, each patient who had complete data would have been weighted by $1/\pi_i$ to account for the patients who have incomplete data. We, therefore, modify the m-th diagonal element of the weight matrix $\mathbf{Q}_{i,1kl}$ as follows:

$$Q_{im,1kl}(\pi_i) = \frac{\Delta_i}{\pi_i} \left[\frac{R_i Z_{ki}}{\eta_k} + \frac{(1 - R_i) Z'_{li}}{\zeta_l} \right], \tag{2.14}$$

and the weight matrix $\mathbf{Q}_{i,1kl}(\pi_i) = \text{diag}(Q_{im,1kl}(\pi_i), m = 1, \ldots, M_i)$. However, π_i is unknown and needs to be estimated. Let $\pi_i = G_i(\boldsymbol{\gamma})$ be the postulated model for drop-out process defined by a set of parameters $\boldsymbol{\gamma}$ and $\hat{\boldsymbol{\gamma}}$ be a regular and asymptotically linear (RAL) estimator of $\boldsymbol{\gamma}$. The weight in (2.14) is then re-defined by replacing π_i by $G_i(\hat{\boldsymbol{\gamma}})$; i.e. the m-th diagonal element of $\boldsymbol{Q}_{i,1kl}(\hat{\boldsymbol{\gamma}})$ is $\{G_i^{-1}(\hat{\boldsymbol{\gamma}}) \Delta_i [\eta_k^{-1} R_i Z_{ki} + \zeta_l^{-1} (1 - R_i) Z'_{li}]\}$. Note that we have

slightly abused the notation to denote this weight matrix by $\boldsymbol{Q}_{i,1kl}(\hat{\boldsymbol{\gamma}})$ instead of $\boldsymbol{Q}_{i,1kl}(\hat{\pi}_i)$. Given a treatment regime $A_1B_kB_l'$, the estimating equation for the IPWGEE is then

$$\sum_{i=1}^{n} \mathbf{U}_i^* \left(\boldsymbol{\beta}_{1kl}, \hat{\boldsymbol{\gamma}}\right) = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \mathbf{Q}_{i,1kl}(\hat{\boldsymbol{\gamma}}) \left\{\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_{1kl}\right\} = \mathbf{0}, \tag{2.15}$$

where $\mathbf{Y}_i$, $\mathbf{X}_i$, and $\mathbf{V}_{i,1kl}$ are defined as in (2.6). Again, as before, the solution of (2.15), $\hat{\boldsymbol{\beta}}_{1kl}^*$, can be obtained through the iterative algorithm.

Lemma 2.4.1. *Under SRA in (2.5), and when $\pi_i = G_i(\boldsymbol{\gamma})$ is known,* $E[\mathbf{Q}_{i,1kl}(\boldsymbol{\gamma}) \mid \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB_l')] = \mathbf{I}_{M_i}$.

Proof. Using iterated conditioning,

$$\begin{aligned}
&E\left[Q_{im,1kl}(\boldsymbol{\gamma}) \mid \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB_l')\right] \\
&= E\left\{E\left[\frac{\Delta_i}{G_i(\boldsymbol{\gamma})} Q_{im,1kl} \mid \Delta_i, R_i, \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB_l')\right]\right\} \\
&= E\left\{\frac{\Delta_i}{G_i(\boldsymbol{\gamma})} E\left[Q_{im,1kl} \mid \Delta_i, R_i, \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB_l')\right]\right\} \\
&= E\left[\frac{\Delta_i}{G_i(\boldsymbol{\gamma})} \times 1\right] \\
&= E\left\{\frac{1}{G_i(\boldsymbol{\gamma})} E\left[\Delta_i \mid R_i, \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB_l')\right]\right\} \\
&= \frac{1}{G_i(\boldsymbol{\gamma})} \times \pi_i = 1, \text{ for } m = 1, \ldots, M_i.
\end{aligned}$$

□

Lemma 2.4.2. *Under CA in (2.4), SRA in (2.5), and when $\pi_i = G_i(\boldsymbol{\gamma})$ is correctly specified,* $E[\mathbf{U}_i^*(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma})] = \mathbf{0}$.

Proof. Under CA in (2.4), $\mathbf{Q}_{i,1kl}(\boldsymbol{\gamma})\mathbf{Y}_i = \mathbf{Q}_{i,1kl}(\boldsymbol{\gamma})\mathbf{Y}_i(A_1B_kB_l')$. Thus,

$$\begin{aligned}
E\left[\mathbf{U}_i^*\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right)\right] &= E\left[\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \mathbf{Q}_{i,1kl}(\boldsymbol{\gamma}) \left\{\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_{1kl}\right\}\right] \\
&= E\left[\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \mathbf{Q}_{i,1kl}(\boldsymbol{\gamma}) \left\{\mathbf{Y}_i(A_1B_kB_l') - \mathbf{X}_i \boldsymbol{\beta}_{1kl}\right\}\right] \quad \text{(by CA in (2.4))} \\
&= E\left\{E\left[\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \mathbf{Q}_{i,1kl}(\boldsymbol{\gamma}) \left\{\mathbf{Y}_i(A_1B_kB_l') - \mathbf{X}_i \boldsymbol{\beta}_{1kl}\right\} \mid \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB_l')\right]\right\} \\
&= E\left[\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \left\{\mathbf{Y}_i(A_1B_kB_l') - \mathbf{X}_i \boldsymbol{\beta}_{1kl}\right\}\right] \quad \text{(by Lemma 2.4.1)} \\
&= E\left\{E\left[\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \left\{\mathbf{Y}_i(A_1B_kB_l') - \mathbf{X}_i \boldsymbol{\beta}_{1kl}\right\} \mid \mathbf{X}_i\right]\right\} \\
&= E\left\{\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \left\{E\left[\mathbf{Y}_i(A_1B_kB_l') \mid \mathbf{X}_i\right] - \mathbf{X}_i \boldsymbol{\beta}_{1kl}\right\}\right\} = \mathbf{0} \quad \text{(by Equation (2.3))}.
\end{aligned}$$

□

Proposition 2.4.1. *$\hat{\boldsymbol{\beta}}^*_{1kl}$ is a consistent estimator of $\boldsymbol{\beta}_{1kl}$, provided (1) $\hat{\boldsymbol{\gamma}}$ is $\sqrt{n}$-consistent, (2) $G_i(\boldsymbol{\gamma})$ is specified correctly, and (3) the probability of having complete data does not depend on counterfactuals.*

Proof. From (1), we can write $\sqrt{n}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}) = n^{-1/2}\sum_{i=1}^n \boldsymbol{\varphi}_i(\boldsymbol{\gamma}) + \mathbf{o_p}(\mathbf{1})$, where $\boldsymbol{\varphi}_i(\boldsymbol{\gamma})$ is the influence function of the estimator $\hat{\boldsymbol{\gamma}}$ and $E[\boldsymbol{\varphi}_i(\boldsymbol{\gamma})] = \mathbf{0}$. Since $\hat{\boldsymbol{\beta}}^*_{1kl}$ satisfies Equation (2.15), the estimator $(\hat{\boldsymbol{\beta}}^*_{1kl}, \hat{\boldsymbol{\gamma}})^T$ is an M-estimator [Stefanski and Boos, 2002] defined by

$$\sum_{i=1}^n \boldsymbol{\Psi}_i(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}) = \sum_{i=1}^n \left\{ \begin{array}{c} \mathbf{U}_i^*(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}) \\ \boldsymbol{\varphi}_i(\boldsymbol{\gamma}) \end{array} \right\} = \mathbf{0}.$$

The consistency of $\hat{\boldsymbol{\beta}}^*_{1kl}$ follows from the fact that $E[\boldsymbol{\Psi}_i(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma})] = \mathbf{0}$ (by Lemma 2.4.2). □

Proposition 2.4.2. *Under assumptions similar to Proposition 2.4.1, $\hat{\boldsymbol{\beta}}^*_{1kl}$ is an asymptotically normally distributed with mean $\boldsymbol{\beta}_{1kl}$ and variance Σ^*/n, where*

$$\left.\begin{array}{l}
\boldsymbol{\Sigma}^* = \mathbf{C}^{*^{-1}}(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma})\, \mathbf{B}^*(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma})\, \mathbf{C}^{*^{-1}}(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma});\\
\mathbf{C}^*(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}) = E\left[\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \mathbf{X}_i\right];\\
\mathbf{B}^*(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}) = E\left[\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \mathbf{Q}_{i,1kl}(\boldsymbol{\gamma}) \{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\} \{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\}^T \mathbf{Q}_{i,1kl}(\boldsymbol{\gamma}) \mathbf{V}_{i,1kl}^{-1} \mathbf{X}_i\right]\\
\qquad -E\left[\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\} \boldsymbol{\varphi}_i^T(\boldsymbol{\gamma})\right] \mathbf{D}^{*^T}(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma})\\
\qquad -\mathbf{D}^*(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}) E\left[\boldsymbol{\varphi}_i(\boldsymbol{\gamma}) \{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\}^T \mathbf{V}_{i,1kl}^{-1} \mathbf{X}_i\right]\\
\qquad +\mathbf{D}^*(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}) E\left[\boldsymbol{\varphi}_i(\boldsymbol{\gamma}) \boldsymbol{\varphi}_i^T(\boldsymbol{\gamma})\right] \mathbf{D}^{*^T}(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma});\\
\mathbf{D}^*(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}) = E\left[\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\} G_i^{-1}(\boldsymbol{\gamma}) \frac{\partial}{\partial \boldsymbol{\gamma}^T} G_i(\boldsymbol{\gamma})\right].
\end{array}\right\} \quad (2.16)$$

Proof. We start with the fact that the estimator $\hat{\boldsymbol{\beta}}^*_{1kl}$ satisfies $\sum_{i=1}^n \mathbf{U}_i^*(\hat{\boldsymbol{\beta}}_{1kl}, \hat{\boldsymbol{\gamma}}) = \mathbf{0}$. First, expanding $\sum_{i=1}^n \mathbf{U}_i^*(\hat{\boldsymbol{\beta}}_{1kl}, \hat{\boldsymbol{\gamma}})$ around $\boldsymbol{\beta}_{1kl}$ using Taylor's expansion, we obtain

$$n^{\frac{1}{2}}\left(\hat{\boldsymbol{\beta}}^*_{1kl} - \boldsymbol{\beta}_{1kl}\right) = \left[-\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}_{1kl}^T} \mathbf{U}_i^*(\boldsymbol{\beta}_{1kl}, \hat{\boldsymbol{\gamma}})\right]^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \mathbf{U}_i^*(\boldsymbol{\beta}_{1kl}, \hat{\boldsymbol{\gamma}}) + \mathbf{o_p}(\mathbf{1}) \quad (2.17)$$

Then, applying Taylor's expansion on (2.17) around $\boldsymbol{\gamma}$, it can be rewritten as

$$\begin{aligned}
n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}^*_{1kl} - \boldsymbol{\beta}_{1kl}) = &\left[-\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}_{1kl}^T} \mathbf{U}_i^*(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma})\right]^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \mathbf{U}_i^*(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma})\\
&+ \left\{\left[-\frac{1}{n}\sum_{i=1}^n \frac{\partial^2}{\partial \boldsymbol{\beta}_{1kl}^T \partial \boldsymbol{\gamma}^T} \mathbf{U}_i^*(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma})\right]^{-1} \frac{1}{n}\sum_{i=1}^n \mathbf{U}_i^*(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma})\right.\\
&\left.+ \left[-\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}_{1kl}^T} \mathbf{U}_i^*(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma})\right]^{-1} \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\gamma}^T} \mathbf{U}_i^*(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma})\right\} \times n^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + \mathbf{o_p}(\mathbf{1}). \quad (2.18)
\end{aligned}$$

Note the convergence of the following quantities in (2.18) as $n \to \infty$:

$$-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial \boldsymbol{\beta}_{1kl}^{T}}\mathbf{U}_i^{*}\left(\boldsymbol{\beta}_{1kl},\boldsymbol{\gamma}\right)\xrightarrow{p} E\left[\mathbf{X}_i^{T}\mathbf{V}_{i,1kl}^{-1}\mathbf{X}_i\right]=\mathbf{C}^{*}\left(\phi,\boldsymbol{\alpha},\boldsymbol{\beta}_{1kl},\boldsymbol{\gamma}\right), \tag{2.19}$$

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{U}_i^{*}\left(\boldsymbol{\beta}_{1kl},\boldsymbol{\gamma}\right)\xrightarrow{p}\mathbf{0}\quad\text{(by Lemma 2.4.2), and} \tag{2.20}$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial \boldsymbol{\gamma}^{T}}\mathbf{U}_i^{*}\left(\boldsymbol{\beta}_{1kl},\boldsymbol{\gamma}\right)\xrightarrow{p}-E\left[\mathbf{X}_i^{T}\mathbf{V}_{i,1kl}^{-1}\left\{\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}_{1kl}\right\}G_i^{-1}(\boldsymbol{\gamma})\frac{\partial}{\partial \boldsymbol{\gamma}^{T}}G_i(\boldsymbol{\gamma})\right]. \tag{2.21}$$

Using (2.19) – (2.21), we can write (2.18) as $n^{1/2}(\hat{\boldsymbol{\beta}}_{1kl}^{*}-\boldsymbol{\beta}_{1kl})=n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\psi}_{i,1kl}^{*}+\mathbf{o_p}(\mathbf{1})$, where $\boldsymbol{\psi}_{i,1kl}^{*}$ is the influence function of the estimator $\hat{\boldsymbol{\beta}}_{1kl}^{*}$ and can be expressed as

$$\mathbf{C}^{*^{-1}}\left(\phi,\boldsymbol{\alpha},\boldsymbol{\beta}_{1kl},\boldsymbol{\gamma}\right)\left\{\mathbf{X}_i^{T}\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}(\boldsymbol{\gamma})\left\{\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}_{1kl}\right\}-\mathbf{D}^{*}\left(\phi,\boldsymbol{\alpha},\boldsymbol{\beta}_{1kl},\boldsymbol{\gamma}\right)\boldsymbol{\varphi}_i(\boldsymbol{\gamma})\right\}, \tag{2.22}$$

where $\mathbf{D}^{*}(\phi,\boldsymbol{\alpha},\boldsymbol{\beta}_{1kl},\boldsymbol{\gamma})=E[\mathbf{X}_i^{T}\mathbf{V}_{i,1kl}^{-1}\{\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}_{1kl}\}G_i^{-1}(\boldsymbol{\gamma})\frac{\partial}{\partial \boldsymbol{\gamma}^{T}}G_i(\boldsymbol{\gamma})]$. By central limit theorem, we can deduce that $n^{1/2}(\hat{\boldsymbol{\beta}}_{1kl}^{*}-\boldsymbol{\beta}_{1kl})\xrightarrow{d}\text{MVN}\left(\mathbf{0},\boldsymbol{\Sigma}^{*}\right)$, where $\boldsymbol{\Sigma}^{*}=E[\boldsymbol{\psi}_{i,1kl}^{*}\boldsymbol{\psi}_{i,1kl}^{*^{T}}]$ and is given in (2.16). □

The asymptotic variance-covariance matrix of $\hat{\boldsymbol{\beta}}_{1kl}^{*}$ can be estimated by the empirical estimator

$$\widehat{\text{var}}\left(\hat{\boldsymbol{\beta}}_{1kl}^{*}\right)=\frac{1}{n}\widehat{E}\left[\boldsymbol{\psi}_{i,1kl}^{*}\boldsymbol{\psi}_{i,1kl}^{*^{T}}\right]=\frac{1}{n^{2}}\sum_{i=1}^{n}\hat{\boldsymbol{\psi}}_{i,1kl}^{*}\hat{\boldsymbol{\psi}}_{i,1kl}^{*^{T}}, \tag{2.23}$$

or by the model-based estimator

$$\widehat{\text{var}}\left(\hat{\boldsymbol{\beta}}_{1kl}^{*}\right)=\frac{1}{n}E\left[\mathbf{C}_n^{*^{-1}}\left(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl}^{*},\hat{\boldsymbol{\gamma}}\right)\mathbf{B}_n^{*}\left(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl}^{*},\hat{\boldsymbol{\gamma}}\right)\mathbf{C}_n^{*^{-1}}\left(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl}^{*},\hat{\boldsymbol{\gamma}}\right)\right], \tag{2.24}$$

where

$$\hat{\boldsymbol{\psi}}_{i,1kl}^{*}=\mathbf{C}_n^{*^{-1}}(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl}^{*},\hat{\boldsymbol{\gamma}})\{\mathbf{X}_i^{T}\hat{\mathbf{V}}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}(\hat{\boldsymbol{\gamma}})[\mathbf{Y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}_{1kl}^{*}]-\mathbf{D}_n^{*}(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl}^{*},\hat{\boldsymbol{\gamma}})\boldsymbol{\varphi}_i(\hat{\boldsymbol{\gamma}})\},$$

$$\mathbf{C}_n^{*}(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl}^{*},\hat{\boldsymbol{\gamma}})=\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i^{T}\hat{\mathbf{V}}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}(\hat{\boldsymbol{\gamma}})\mathbf{X}_i,$$

$$\mathbf{D}_n^{*}(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl}^{*},\hat{\boldsymbol{\gamma}})=\frac{1}{n}\sum_{i=1}^{n}[\mathbf{X}_i^{T}\hat{\mathbf{V}}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}(\hat{\boldsymbol{\gamma}})\{\mathbf{Y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}_{1kl}^{*}\}G_i^{-1}(\hat{\boldsymbol{\gamma}})\dot{G}_i(\hat{\boldsymbol{\gamma}})],$$

and

$$
\begin{aligned}
&\mathbf{B}_n^*\left(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl}^*,\hat{\boldsymbol{\gamma}}\right)\\
=\;&\frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{X}_i^T\hat{\mathbf{V}}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}(\hat{\boldsymbol{\gamma}})\left\{\mathbf{Y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}_{1kl}^*\right\}\left\{\mathbf{Y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}_{1kl}^*\right\}^T\mathbf{Q}_{i,1kl}(\hat{\boldsymbol{\gamma}})\hat{\mathbf{V}}_{i,1kl}^{-1}\mathbf{X}_i\right]\\
&-\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{X}_i^T\hat{\mathbf{V}}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}(\hat{\boldsymbol{\gamma}})\left\{\mathbf{Y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}_{1kl}^*\right\}\boldsymbol{\varphi}_i^T(\hat{\boldsymbol{\gamma}})\right]\right\}\times\mathbf{D}_n^*\left(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl}^{*^T},\hat{\boldsymbol{\gamma}}\right)\\
&-\mathbf{D}_n^*\left(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl}^*,\hat{\boldsymbol{\gamma}}\right)\times\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\boldsymbol{\varphi}_i(\hat{\boldsymbol{\gamma}})\left\{\mathbf{Y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}}_{1kl}^*\right\}^T\mathbf{Q}_{i,1kl}(\hat{\boldsymbol{\gamma}})\hat{\mathbf{V}}_{i,1kl}^{-1}\mathbf{X}_i\right]\right\}\\
&+\mathbf{D}_n^*\left(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl}^*,\hat{\boldsymbol{\gamma}}\right)\times\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\boldsymbol{\varphi}_i(\hat{\boldsymbol{\gamma}})\boldsymbol{\varphi}_i^T(\hat{\boldsymbol{\gamma}})\right]\right\}\times\mathbf{D}_n^{*^T}\left(\hat{\phi},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{1kl}^*,\hat{\boldsymbol{\gamma}}\right).
\end{aligned}
$$

The estimators of ϕ, $\boldsymbol{\alpha}$, $\mathbf{V}_{i,1kl}$, and $\mathbf{A}_{i,1kl}$ are obtained by replacing $\hat{\boldsymbol{\beta}}_{1kl}$ with $\hat{\boldsymbol{\beta}}_{1kl}^*$ and $\mathbf{Q}_{i,1kl}$ with $\mathbf{Q}_{i,1kl}(\hat{\boldsymbol{\gamma}})$ in (2.12) and (2.13).

2.5 COMPARISON AMONG TREATMENT REGIMES

To compare the effects of various treatment regimes, contrasts of target regime effects can be constructed and tested via Wald test. Since a patient can belong to more than one regime, the estimators of treatment effects among regimes will be correlated. In such situations, the covariance between estimators needs to be estimated. In the case of four treatment regimes, $A_1B_kB_l'$ for $k,l\in\{1,2\}$, one can consider six pairwise comparisons of the estimators of these four treatment regimes. Among these six pairs of the estimators, the covariance of $(\hat{\boldsymbol{\beta}}_{111}^*,\hat{\boldsymbol{\beta}}_{122}^*)$ will be zero, since $\hat{\boldsymbol{\beta}}_{111}^*$ and $\hat{\boldsymbol{\beta}}_{122}^*$ are estimated using data from two different subgroups of patients. For the same reason, the covariance of $(\hat{\boldsymbol{\beta}}_{112}^*,\hat{\boldsymbol{\beta}}_{121}^*)$ will also be zero. However, the estimators $\hat{\boldsymbol{\beta}}_{111}^*$ and $\hat{\boldsymbol{\beta}}_{112}^*$ are correlated, since both estimators used the same information from those patients who received A_1 and followed with B_1 after responding to A_1. Hence, the covariance of $(\hat{\boldsymbol{\beta}}_{111}^*,\hat{\boldsymbol{\beta}}_{112}^*)$ needs to be estimated. Similarly, we will have to estimate the covariances for the pairs $(\hat{\boldsymbol{\beta}}_{111}^*,\hat{\boldsymbol{\beta}}_{121}^*)$, $(\hat{\boldsymbol{\beta}}_{112}^*,\hat{\boldsymbol{\beta}}_{122}^*)$, and $(\hat{\boldsymbol{\beta}}_{121}^*,\hat{\boldsymbol{\beta}}_{122}^*)$.

To estimate the covariance between two correlated estimators, we use the fact that the large-sample covariance between two estimators can be obtained through the expectation of the product of their influence functions. We will demonstrate covariance computation

for the pair of estimators $\hat{\boldsymbol{\beta}}^*_{111}$ and $\hat{\boldsymbol{\beta}}^*_{112}$. Similar computation follows for other correlated pairs. The covariance of $\hat{\boldsymbol{\beta}}^*_{111}$ and $\hat{\boldsymbol{\beta}}^*_{112}$ is given by $n^{-1}E[\boldsymbol{\psi}^*_{i,111}\boldsymbol{\psi}^{*^T}_{i,112}]$, where $\boldsymbol{\psi}^*_{i,111}$ and $\boldsymbol{\psi}^{*^T}_{i,112}$ are the influence functions of estimators $\hat{\boldsymbol{\beta}}^*_{111}$ and $\hat{\boldsymbol{\beta}}^*_{112}$, which are obtained from Equation (2.22). Therefore, one can estimate the covariance of $\hat{\boldsymbol{\beta}}^*_{111}$ and $\hat{\boldsymbol{\beta}}^*_{112}$ in a manner similar to the estimation of variance-covariance matrix through Equation (2.23) or (2.24).

To test the effects of treatment regimes $A_1B_kB'_l$ for $k, l \in \{1, 2\}$, $\boldsymbol{\beta} = [\boldsymbol{\beta}^T_{111}, \boldsymbol{\beta}^T_{112}, \boldsymbol{\beta}^T_{121}, \boldsymbol{\beta}^T_{122}]^T$ via Wald test, one can establish the null hypothesis of $\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$, where each $\boldsymbol{\beta}_{1kl}$ is a $(p+2)$-dimensional vector and $\mathbf{A}$ is a matrix with $[4(p+2)]$ columns such that $\text{rank}(A) < [4(p+2)]$. The test statistic will be $T = (\mathbf{A}\hat{\boldsymbol{\beta}})^T(\mathbf{A}\hat{\boldsymbol{\Sigma}}\mathbf{A}^T)^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}}) \sim \chi^2_{\text{rank}(A)}$, where $\hat{\boldsymbol{\beta}} = [\hat{\boldsymbol{\beta}}^{*^T}_{111}, \hat{\boldsymbol{\beta}}^{*^T}_{112}, \hat{\boldsymbol{\beta}}^{*^T}_{121}, \hat{\boldsymbol{\beta}}^{*^T}_{122}]^T$ and $\hat{\boldsymbol{\Sigma}}$ is the estimated covariance matrix of $\hat{\boldsymbol{\beta}}$. Each element of $\hat{\boldsymbol{\Sigma}}$ can be obtained by calculating the covariances of all pairs of estimators.

2.6 SIMULATION STUDY

To evaluate the performance of the estimators in small samples, we conducted several simulation studies. We simulated data from a population which has similar design to the REVAMP study. Each patient will have repeated measurements at each visit so that $\mathbf{t}_i = (0, 2, 4, 6, 8, 10, 12)^T$ for patient i. We considered one initial treatment A_1, two second stage treatments, B_1 and B_2, for responders, and two second stage treatments, B'_1 and B'_2, for non-responders. All patients received initial treatment A_1 at $t_{i1} = 0$. For each patient i, we generated counterfactual random variables listed in (2.1). At $t_{i3} = 4$ and $t_{i4} = 6$, patients were assessed to see if they had responded to the initial treatment A_1. Thus, the index of the time of response in this case is $m_{1i} = 3$ or 4. The response status, R_i, was drawn from a Bernoulli(θ) distribution. We assume that of these responders, $100 \times \theta_1\%$ responded at time t_{i3} and the rest at time t_{i4}. Thus each responder was assigned randomly to respond at time t_{i3} with probability θ_1. We considered two sets of parameters, $(\theta, \theta_1) = (0.5, 0.25)$ and $(0.3, 0.15)$ meaning 50% (30%) of the patients would respond of which 25% (15%) would respond at time t_{i3} and the rest at time t_{i4}. We also generated a baseline covariate, age, which follows a normal distribution with mean $\mu_{age} = 45$ and standard deviation $\sigma_{age} = 11$.

For each patient i, counterfactual outcome vectors $\mathbf{Y}_i(A_1)$, $\mathbf{Y}_i(A_1B_1)$, $\mathbf{Y}_i(A_1B_2)$, $\mathbf{Y}_i(A_2B_1)$, and $\mathbf{Y}_i(A_2B_2')$, were generated from the following multivariate noraml distributions (MVN):

$$\begin{aligned}
\mathbf{Y}_i(A_1) &\sim \text{MVN}_{m_{1i}}((\theta_{1,A_1} + \theta_{2,A_1} \times \text{age}_i) \times \mathbf{1}_{m_{1i}} + \theta_{3,A_1} \times \mathbf{t}_{0i}, \mathbf{\Sigma}_0),\\
\mathbf{Y}_i(A_1B_1) &\sim \text{MVN}_{7-m_{1i}}((\theta_{1,A_1B_1} + \theta_{2,A_1B_1} \times \text{age}_i) \times \mathbf{1}_{(7-m_{1i})} + \theta_{3,A_1B_1} \times \mathbf{t}_{1i}, \mathbf{\Sigma}_1),\\
\mathbf{Y}_i(A_1B_2) &\sim \text{MVN}_{7-m_{1i}}((\theta_{1,A_1B_2} + \theta_{2,A_1B_2} \times \text{age}_i) \times \mathbf{1}_{(7-m_{1i})} + \theta_{3,A_1B_2} \times \mathbf{t}_{1i}, \mathbf{\Sigma}_1),\\
\mathbf{Y}_i(A_1B_1') &\sim \text{MVN}_{7-m_{1i}}((\theta_{1,A_1B_1'} + \theta_{2,A_1B_1'} \times \text{age}_i) \times \mathbf{1}_{(7-m_{1i})} + \theta_{3,A_1B_1'} \times \mathbf{t}_{1i}, \mathbf{\Sigma}_1),\\
\text{and } \mathbf{Y}_i(A_1B_2') &\sim \text{MVN}_{7-m_{1i}}((\theta_{1,A_1B_2'} + \theta_{2,A_1B_2'} \times \text{age}_i) \times \mathbf{1}_{(7-m_{1i})} + \theta_{3,A_1B_2'} \times \mathbf{t}_{1i}, \mathbf{\Sigma}_1),
\end{aligned}$$

where m_{1i} equals to either 3 or 4; $\mathbf{t}_{0i}$ is a $m_{1i}\times 1$ vector which indicates weeks of measurements in the first stage and $\mathbf{t}_{1i}$ is a $(7-m_{1i})\times 1$ vector which indicates weeks of measurements in the second stage; $\mathbf{\Sigma}_0(\sigma_0, \rho_0)$ is a $m_{1i}\times m_{1i}$ covariance matrix defined by standard deviation σ_0 and autoregressive correlation ρ_0; $\mathbf{\Sigma}_1(\sigma_1, \rho_1)$ is a $(7-m_{1i})\times(7-m_{1i})$ covariance matrix defined by standard deviation σ_1 and autoregressive correlation ρ_1. The values of σ_0, σ_1, ρ_0, and ρ_1 were set to be 5, 3, 0.8, and 0.8, respectively. For each patient i, the m-th element of the counterfactual outcome under a treatment regime $A_1B_kB_l'$ for $k, l = 1, 2$ and $m = 1, 2, \ldots, 7$ is generated using Equation (2.2). The estimation of parameters $\boldsymbol{\beta}_{1kl} = (\beta_{0,1kl}, \beta_{1,1kl}, \beta_{2,1kl})^T$ in Equation (2.3) of the model $E[Y_{im}(A_1B_kB_l') \mid \mathbf{x}_{im}] = \mathbf{x}_{im}^T\boldsymbol{\beta}_{1kl}$ was the main focus in this article, where $\mathbf{x}_{im} = [1, t_{im}, \text{age}_i]^T$.

We considered the following parameter values for each counterfactual vector: θ_{1,A_1}= 25, θ_{2,A_1}= 0.5, θ_{3,A_1}= -0.5, θ_{1,A_1B_1}= 27, θ_{2,A_1B_1}= 0.6, θ_{3,A_1B_1}= -1.5, θ_{1,A_1B_2}= 38, θ_{2,A_1B_2}= 0.4, θ_{3,A_1B_2}= -2, $\theta_{1,A_1B_1'}$= 36, $\theta_{2,A_1B_1'}$= 0.7, $\theta_{3,A_1B_1'}$= -3, $\theta_{1,A_1B_2'}$= 68, $\theta_{2,A_1B_2'}$= 0.3, and $\theta_{3,A_1B_2'}$= -5. For this population, the true parameter values were $\boldsymbol{\beta}_{111}^T = [23.09, -1.13, 0.57]$, $\boldsymbol{\beta}_{112}^T = [27.94, -1.53, 0.48]$, $\boldsymbol{\beta}_{121}^T = [25.45, -1.30, 0.52]$, and $\boldsymbol{\beta}_{122}^T = [30.38, -1.69, 0.43]$ for 50% response, and $\boldsymbol{\beta}_{111}^T = [22.88, -1.20, 0.58]$, $\boldsymbol{\beta}_{112}^T = [29.65, -1.77, 0.45]$, $\boldsymbol{\beta}_{121}^T = [24.28, -1.31, 0.55]$, and $\boldsymbol{\beta}_{122}^T = [31.10, -1.86, 0.42]$ for 30% response.

We simulated 2000 Monte Carlo samples of sizes 250 and 400 observations from the populations described above with the following characteristics. For responders (i.e. $R_i = 1$), the assignment indicator, Z_{1i}, for treatment B_1 was drawn from a Bernoulli distribution with probability η_1; the assignment indicator for treatment B_2 was defined as $Z_{2i} = 1 - Z_{1i}$. For

non-responders (i.e. $R_i = 0$), the assignment indicator, Z'_{1i}, for treatment B'_1 was drawn from a Bernoulli distribution with probability ζ_1; the assignment indicator for treatment B'_2 was defined as $Z'_{2i} = 1 - Z'_{1i}$. We defined the observed outcome for patient i at t_{im} as in Equation (2.4).

Additionally, a number of patients were allowed to drop out for the purpose of illustration based on the following logistic regression model: $\pi_i = G_i(\boldsymbol{\gamma}) = \Pr(\Delta_i = 1 \mid \boldsymbol{\gamma}, \text{age}, Y_{i1}) = [1 + \exp(-\gamma_0 - \gamma_1 \times \text{age}_i - \gamma_2 \times Y_{i1})]^{-1}$, where $\boldsymbol{\gamma} = [\gamma_0, \gamma_1, \gamma_2]^T$; Y_{i1} is the outcome at $t_{i1} = 0$ for patient i. Choices of values of parameters $\boldsymbol{\gamma} = [5.4, -0.02, -0.07]^T$ and $\boldsymbol{\gamma} = [5.1, -0.03, -0.08]^T$ gave us approximate drop-out rates of 25% and 50%, respectively. For sample sizes of 250 and 400 and response rates of 50% and 30%, we considered the following drop-out rates and analysis strategies: (1) no drop-out with weighting for randomization ($\mathbf{Q}$); (2) 25% drop-out rate with weighting for randomization only ($\mathbf{Q}$); (3) 25% drop-out rate with weighting for randomization and drop-out ($\mathbf{Q}(\boldsymbol{\gamma})$); (4) 50% drop-out rate with weighting for randomization only ($\mathbf{Q}$); (5) 50% drop-out rate with weighting for randomization and drop-out ($\mathbf{Q}(\boldsymbol{\gamma})$).

Table 2.1 shows the simulation results of estimation for samples of size 250. When the missing was due to randomization only (i.e. no drop-out), the IPWGEE estimators with $\mathbf{Q}$ as weight matrix were approximately unbiased for all regimes regardless of the response rates. The maximum relative bias observed was 0.8%. The estimated standard deviations of the estimators were consistent with the Monte Carlo standard deviations. The coverage probabilities for the 95% Wald confidence intervals for the parameters were between 92.4% and 95.7%. For the 25% drop-out rate, the estimators were biased when the drop-out were ignored (i.e. used $\mathbf{Q}$ as weight matrix in the IPWGEE.) The relative biases ranged from 0.8 to 3.4%. When the analysis accounted for drop-outs using $\mathbf{Q}(\boldsymbol{\gamma})$ as weight matrix in the IPWGEE, the estimators were approximately unbiased. For example, the estimator weighted by $\mathbf{Q}$ for regime $A_1B_2B'_2$ with 30% response rate had relative bias 3.2%, however, the estimator weighted by $\mathbf{Q}(\boldsymbol{\gamma})$ for the same had near zero bias. The estimated standard deviations of estimators weighted by $\mathbf{Q}(\boldsymbol{\gamma})$ were consistent with corresponding Monte Carlo standard deviations. The Wald confidence intervals achieved coverage close to their nominal confidence levels. When the drop-out rate was raised from 25% to 50%, similar results were

observed. The estimators weighted by $\mathbf{Q}(\boldsymbol{\gamma})$ remained approximately unbiased for all regimes under both 50% and 30% response rates. Standard errors of estimators were slightly higher for higher drop-out rate.

For $n = 400$ (shown in Table 2.2), the estimators with proper weighting were approximately unbiased regardless of response and drop-out rates. The standard errors of estimators were, as expected, smaller than those for $n = 250$. The coverage probabilities for the 95% Wald confidence intervals ranged from 92.4% to 95.2%.

2.7 ANALYSIS OF THE REVAMP DATA

A total of 808 patients with chronic forms of major depressive disorders were enrolled in the REVAMP study [Trivedi et al., 2008]. Patients were evaluated in two 12-week stages for a maximum of 24 weeks. In the first stage, patients were treated with one of four antidepressants: Sertaline (SERT), Escitalopram (EcCIT), Bupropion (BUP-SR), and Venlafaxine (VLF-XR). The choice of antidepressants was based on the algorithm using information on pharmacotherapy treatment history (see Trivedi et al., 2008 for more details.) During weeks 8 through 12 in the first stage, each patient was evaluated to assess the response. Patients with full response continued their antidepressant in the second stage. Patients who did not meet the criteria of full response were randomly assigned to three groups: Cognitive Behavioral Analysis System of Psychotherapy (CBASP) plus medication, Brief Supportive Psychotherapy (BSP) plus medication, or Medication alone (MED). The randomization was done with unequal probabilities of 0.4, 0.4, and 0.2, respectively, to CBASP, BSP, and MED. For the purpose of illustration, we considered only patients treated with SERT in this article. Hence, we were interested in comparing three treatment regimes: (1) treat with SERT, continue SERT if respond, otherwise add CBASP to SERT; (2) treat with SERT, continue SERT if respond, otherwise add BSP to SERT; and (3) treat with SERT, continue SERT if respond, otherwise add MED to SERT. These three regimes are denoted by SSC, SSB, and SSM, respectively. The response to therapy was measured by the 24-item Hamilton Rating Scale for Depression (HRSD) score at each visit. Figure 2.2 shows the design and

patient flow of 618 patients in the REVAMP study who received SERT. The responses of 136 patients could not be assessed since they dropped out from the study in the first stage. For 482 patients entering the second stage, about 65% of patients completed the study. To show a snapshot of the data, we have presented the HRSD scores of eight selected patients in Figure 2.3. The left panel of Figure 2.3 represents the HRSD scores for patients who had their response status confirmed at various time points and went on to complete the study. The right panel of Figure 2.3 represents the HRSD scores for patients who dropped out from the study (their response statuses may or may not have been confirmed.) Figure 2.3 illustrates how information collected in a two-stage longitudinal study can vary across patients due to response and drop-out. Because of the presence of drop-outs, we applied the IPWGEE method described in Section 2.4 to estimate the effects of three depression treatment regimes in the REVAMP study. The effect of the r-th treatment regime was formulated as the coefficients $\boldsymbol{\beta}_r$ in the marginal mean model:

$$\begin{aligned} E\left[\text{HRSD}_{im}(r) \mid \mathbf{x}_{im}\right] &= \mathbf{x}_{im}^T \boldsymbol{\beta} \\ &= \beta_{0,r} + \beta_{1,r} \times t_{im} + \beta_{2,r} \times \text{age}_i + \beta_{3,r} \times \text{emp}_i + \beta_{4,r} \times \text{age}_i \times t_{im}, \end{aligned}$$

where $r \in \{SSC, SSB, SSM\}$; emp_i=1 if employed and emp_i=0 otherwise. To account for incompleteness through the IPWGEE, the probability of having complete data for patient i, $\pi_i = P\{\Delta_i = 1\} = G_i(\boldsymbol{\gamma})$, was estimated from the data. Table 2.3 shows the distributions of drop-outs across baseline covariates. In univariate analysis using chi-square test (or Fisher's exact test) with level of 0.05, Caucasian race, higher education, and older age were significantly associated with having complete data. Based on the results in Table 2.3 and model selection strategy, a logistic model for π_i was postulated with covariates including Caucasian race, employment status, education level, age, and the baseline observed HRSD score. Specifically,

$$\begin{aligned} \hat{\pi}_i &= G_i\left(\hat{\boldsymbol{\gamma}}\right) \\ &= \left[1 + \exp(-\hat{\gamma}_0 - \hat{\gamma}_1 \times \text{race}_i - \hat{\gamma}_2 \times \text{emp}_i - \hat{\gamma}_3 \times \text{edu}_i - \hat{\gamma}_4 \times \text{age}_i - \hat{\gamma}_5 \times \text{HRSD}_{i1})\right]^{-1}, \end{aligned}$$

where race_i=1 if Caucasian and race_i=0 if otherwise; emp_i=1 if employed and emp_i=0 if otherwise; edu_i=1 if education higher than high school and edu_i=0 if otherwise; HRSD_{i1} is the observed HRSD score at baseline.

The results of the REVAMP data analysis are shown in Table 2.4. Controlling for the baseline employment status, for a patient of age 43 years, the treatment regime SSB will have the highest reduction of 0.741 per week in the HRSD scores from baseline, followed by the treatment regime SSC (0.707/week) and then by the treatment regime SSM (0.702/week). While the effect of each treatment regime was statistically significant ($p<0.001$), the magnitude of the effect was similar across three regimes. In Table 2.5, a Wald Chi-square test comparing the effects of treatment regimes ($H_0 : \beta_{1,SSC}+\beta_{4,SSC}\times 43 = \beta_{1,SSB}+\beta_{4,SSB}\times 43 = \beta_{1,SSM} + \beta_{4,SSM} \times 43$) with 2 degrees of freedom resulted in a p-value of 0.842, indicating that there was no evidence that the effects of these three treatment regimes were significantly different from each other.

2.8 DISCUSSION

In a two-stage longitudinal study, such as the one presented here, drop-out is a common phenomenon. If the drop-out occurs prior to the second randomization, a patient's response status will be unknown. Additionally, in the second stage of the study, patients randomized to one treatment can not receive other competing treatments, hence the fundamental problem of causal inference [Holland, 1986] applies. We have used the inverse-probability-weighted generalized estimating equations (IPWGEE) method (Robins et al., 1995) to take into account the missing data due to randomization and drop-out. The weights are formed by inversely weighting the probability of randomization to the treatment dictated by the regime and the probability of having complete data. The probability of having complete data is estimated through a logit model. We showed that the IPWGEE estimators are consistent and asymptotically normal. We have provided evidence of the bias incurred when appropriate weighting is not applied. We also showed how to compare treatment regimes via

Wald test, which required computation of covariance between two estimated regime effects. We have demonstrated our methods using a dataset from a depression study.

Our methods account for missing data due to randomization and drop-out. However, they ignore the time to response and the time to drop-out in the second stage. It is possible to increase precision of the estimators by incorporating these two phenomenons into the estimation process. We will present this generalization in Chapter 3.

2.9 FIGURES AND TABLES

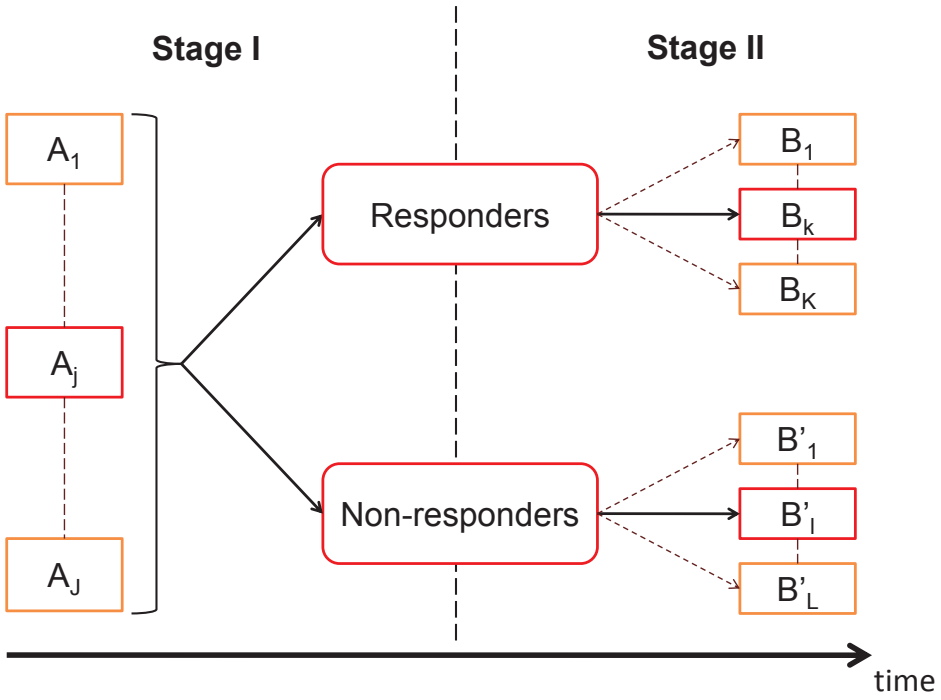


Figure 2.1: **A study with two-stage design.**

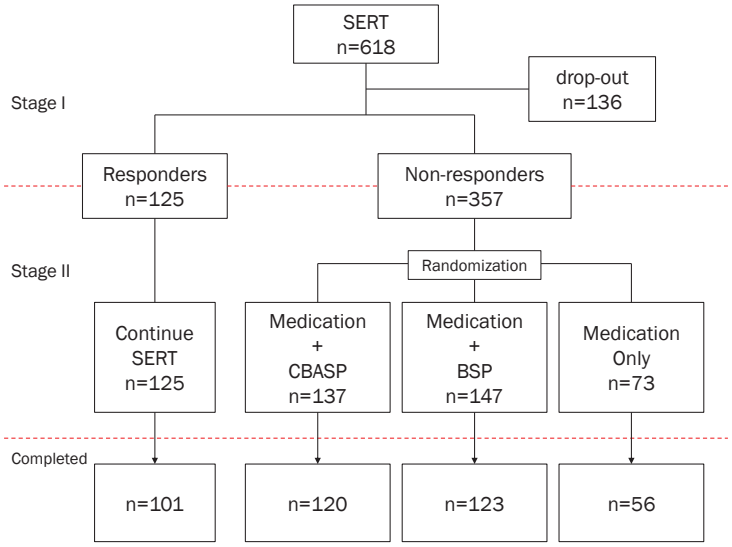


Figure 2.2: **Patient flow in the REVAMP study.**

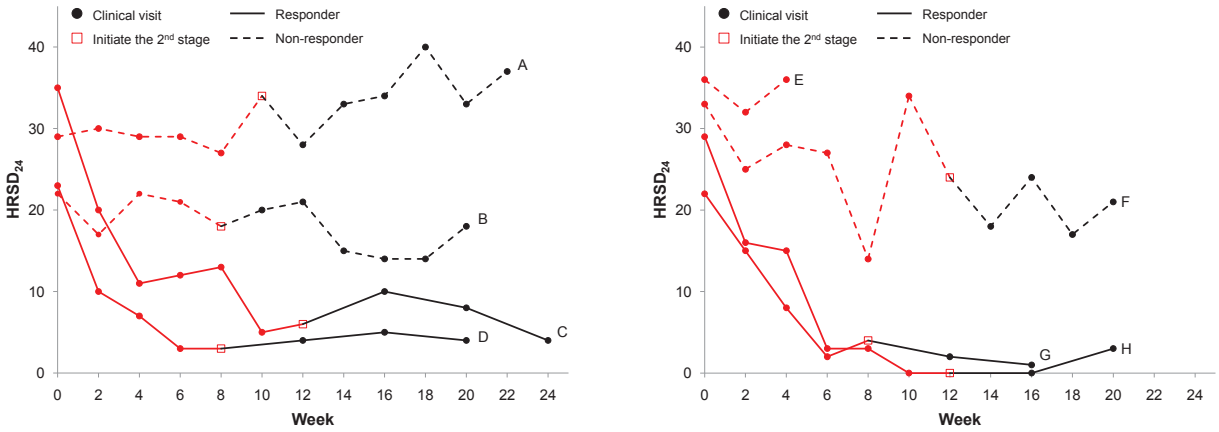


Figure 2.3: **The HRSD scores of eight selected patients in the REVAMP study.** Patients A, B, C, and D completed the study while patients E, F, G, and H did not. Patients A and B are non-responders whose response status was confirmed at weeks 10 and 8, respectively. Patients C and D are responders whose response status was confirmed at weeks 12 and 8, respectively. Patient E dropped out at week 4 without response information. Patient F dropped out at week 20 after failing to respond by week 12. Patients G and H dropped out at weeks 16 and 20 after responding to initial treatment at weeks 8 and 12, respectively. The solid lines represent responders, the dashed lines represent non-responders, and the dotted line represents patient who dropped out before the randomization and his/her response status could not be ascertained. The closed circles show when the HRSD scores were measured, the open squares show when the response status was confirmed.

Table 2.1: **Simulation results of $\hat{\beta}_{1,1kl}$ based on 2000 Monte Carlo (MC) samples of size 250.** EST is MC mean of estimates, SE is MC mean of estimated standard errors, MCSE is standard error of MC estimates, and CP is empirical coverage probability.

Drop-out Rate	Response Rate	Regime	True Value	EST (SE)	MCSE	CP%	EST (SE)	MCSE	CP%
				Weighted by $\mathbf{Q}$			Weighted by $\mathbf{Q}(\boldsymbol{\gamma})$		
0%	50%	$A_1B_1B_1'$	-1.13	-1.13 (0.050)	0.050	95.7	Not applicable		
		$A_1B_1B_2'$	-1.53	-1.53 (0.070)	0.075	94.1	Not applicable		
		$A_1B_2B_1'$	-1.30	-1.30 (0.047)	0.049	92.8	Not applicable		
		$A_1B_2B_2'$	-1.69	-1.70 (0.060)	0.065	92.4	Not applicable		
	30%	$A_1B_1B_1'$	-1.20	-1.20 (0.050)	0.051	93.8	Not applicable		
		$A_1B_1B_2'$	-1.77	-1.76 (0.067)	0.069	94.3	Not applicable		
		$A_1B_2B_1'$	-1.31	-1.30 (0.047)	0.047	95.2	Not applicable		
		$A_1B_2B_2'$	-1.86	-1.86 (0.058)	0.060	94.6	Not applicable		
25%	50%	$A_1B_1B_1'$	-1.13	-1.12 (0.058)	0.060	92.6	-1.13 (0.056)	0.054	96.2
		$A_1B_1B_2'$	-1.53	-1.49 (0.079)	0.084	89.9	-1.53 (0.082)	0.088	95.9
		$A_1B_2B_1'$	-1.30	-1.27 (0.054)	0.057	90.6	-1.31 (0.053)	0.056	92.1
		$A_1B_2B_2'$	-1.69	-1.64 (0.068)	0.072	86.1	-1.71 (0.068)	0.070	94.4
	30%	$A_1B_1B_1'$	-1.20	-1.19 (0.058)	0.058	93.9	-1.21 (0.056)	0.057	94.0
		$A_1B_1B_2'$	-1.77	-1.71 (0.075)	0.079	87.4	-1.76 (0.079)	0.084	93.4
		$A_1B_2B_1'$	-1.31	-1.28 (0.055)	0.056	91.7	-1.30 (0.054)	0.053	94.3
		$A_1B_2B_2'$	-1.86	-1.80 (0.066)	0.069	82.2	-1.86 (0.067)	0.072	93.4
50%	50%	$A_1B_1B_1'$	-1.13	-1.10 (0.071)	0.070	91.8	-1.13 (0.069)	0.076	92.2
		$A_1B_1B_2'$	-1.53	-1.44 (0.093)	0.093	86.2	-1.53 (0.110)	0.117	92.4
		$A_1B_2B_1'$	-1.30	-1.24 (0.067)	0.072	83.4	-1.29 (0.068)	0.079	93.1
		$A_1B_2B_2'$	-1.69	-1.58 (0.081)	0.087	66.8	-1.69 (0.090)	0.100	92.6
	30%	$A_1B_1B_1'$	-1.20	-1.18 (0.072)	0.073	93.4	-1.20 (0.070)	0.075	93.5
		$A_1B_1B_2'$	-1.77	-1.65 (0.090)	0.091	75.3	-1.76 (0.105)	0.110	92.6
		$A_1B_2B_1'$	-1.31	-1.26 (0.068)	0.068	90.1	-1.30 (0.068)	0.070	94.0
		$A_1B_2B_2'$	-1.86	-1.73 (0.080)	0.083	58.2	-1.86 (0.088)	0.095	92.8

Table 2.2: **Simulation results of $\hat{\beta}_{1,1kl}$ based on 2000 Monte Carlo (MC) samples of size 400.** EST is MC mean of estimates, SE is MC mean of estimated standard errors, MCSE is standard error of MC estimates, and CP is empirical coverage probability.

Drop-out Rate	Response Rate	Regime	True Value	EST (SE)	MCSE	CP%	EST (SE)	MCSE	CP%
				Weighted by $\mathbf{Q}$			Weighted by $\mathbf{Q}(\boldsymbol{\gamma})$		
0%	50%	$A_1B_1B_1'$	-1.13	-1.13 (0.042)	0.042	95.5	Not applicable		
		$A_1B_1B_2'$	-1.53	-1.53 (0.060)	0.062	94.4	Not applicable		
		$A_1B_2B_1'$	-1.30	-1.30 (0.040)	0.041	93.9	Not applicable		
		$A_1B_2B_2'$	-1.69	-1.70 (0.051)	0.053	94.7	Not applicable		
	30%	$A_1B_1B_1'$	-1.20	-1.20 (0.039)	0.040	94.5	Not applicable		
		$A_1B_1B_2'$	-1.77	-1.76 (0.053)	0.055	94.1	Not applicable		
		$A_1B_2B_1'$	-1.31	-1.30 (0.038)	0.039	93.1	Not applicable		
		$A_1B_2B_2'$	-1.86	-1.86 (0.046)	0.048	94.6	Not applicable		
25%	50%	$A_1B_1B_1'$	-1.13	-1.12 (0.045)	0.046	94.2	-1.13 (0.044)	0.045	92.4
		$A_1B_1B_2'$	-1.53	-1.49 (0.062)	0.064	89.0	-1.53 (0.065)	0.065	95.2
		$A_1B_2B_1'$	-1.30	-1.27 (0.043)	0.042	90.0	-1.30 (0.042)	0.044	94.5
		$A_1B_2B_2'$	-1.69	-1.64 (0.054)	0.054	80.7	-1.69 (0.055)	0.057	94.6
	30%	$A_1B_1B_1'$	-1.20	-1.19 (0.046)	0.046	94.3	-1.20 (0.044)	0.044	94.2
		$A_1B_1B_2'$	-1.77	-1.71 (0.060)	0.062	83.0	-1.76 (0.063)	0.064	94.6
		$A_1B_2B_1'$	-1.31	-1.29 (0.043)	0.044	91.4	-1.30 (0.042)	0.042	94.6
		$A_1B_2B_2'$	-1.86	-1.80 (0.052)	0.056	76.6	-1.86 (0.053)	0.053	95.0
50%	50%	$A_1B_1B_1'$	-1.13	-1.11 (0.056)	0.057	92.5	-1.13 (0.055)	0.058	93.5
		$A_1B_1B_2'$	-1.53	-1.44 (0.074)	0.074	77.6	-1.53 (0.088)	0.093	93.6
		$A_1B_2B_1'$	-1.30	-1.24 (0.052)	0.056	79.1	-1.30 (0.054)	0.059	93.4
		$A_1B_2B_2'$	-1.69	-1.57 (0.065)	0.066	48.8	-1.70 (0.073)	0.076	93.9
	30%	$A_1B_1B_1'$	-1.20	-1.18 (0.056)	0.058	93.1	-1.20 (0.056)	0.056	94.8
		$A_1B_1B_2'$	-1.77	-1.65 (0.071)	0.072	60.8	-1.76 (0.084)	0.089	94.5
		$A_1B_2B_1'$	-1.31	-1.27 (0.053)	0.054	87.8	-1.30 (0.054)	0.056	93.3
		$A_1B_2B_2'$	-1.86	-1.73 (0.063)	0.063	39.7	-1.86 (0.070)	0.073	93.8

Table 2.3: **Drop-out rates by baseline characteristics.**

Characteristics	Complete (n=400)	Incomplete (n=218)	p-value
Categorical: No. (%)			
Sex			
Male	185 (66.1)	95 (33.9)	0.52
Female	215 (63.6)	123 (36.4)	
Caucasian Race			
Yes	344 (66.9)	170 (33.1)	0.01
No	56 (53.9)	48 (46.2)	
Hispanic ethnicity			
Yes	33 (58.9)	23 (41.1)	0.34
No	362 (65.3)	192 (34.7)	
Missing, No.	5	3	
Employment			
Employed	255 (66.9)	126 (33.1)	0.12
Unemployed	141 (60.8)	91 (39.2)	
Missing, No.	4	1	
Education			
$\leq$ High school	139 (53.9)	119 (46.1)	<0.01
$>$ High school	241 (71.5)	96 (28.5)	
Missing, No.	20	3	
Marital status			
Married	158 (68.7)	72 (31.3)	0.10
Not married	238 (62.1)	145 (37.9)	
Missing, No.	4	1	
Continuous: mean (s.d.)			
Age, year			
Mean (SD)	45.6 (12.2)	39.4 (12.0)	<0.01
Missing, No.	1	0	

Table 2.4: **Estimated effects of treatment regimes in the REVAMP study.**

Regime	Effect	Estimated Parameter	Standard Error	p-value
SSC [1]	Intercept	26.817	2.218	<0.001
	Time	-0.836	0.137	<0.001
	Age	-0.008	0.041	0.842
	Employment	-1.056	1.249	0.398
	Time×Age	0.003	0.003	0.338
SSB [2]	Intercept	25.380	1.659	<0.001
	Time	-0.913	0.076	<0.001
	Age	0.016	0.032	0.632
	Employment	-2.244	0.835	0.007
	Time×Age	0.004	0.002	0.013
SSM [3]	Intercept	29.268	2.821	<0.001
	Time	-0.711	0.116	<0.001
	Age	-0.065	0.048	0.175
	Employment	-3.135	1.423	0.028
	Time×Age	0.0002	0.003	0.939

[1] Treat with SERT, continue SERT if respond, otherwise add CBASP to SERT
[2] Treat with SERT, continue SERT if respond, otherwise add BSP to SERT
[3] Treat with SERT, continue SERT if respond, otherwise add MED to SERT

Table 2.5: **Comparisons of the estimated effects of treatment regimes on the HRSD score for a patient of 43 years controlling for baseline employment status in the REVAMP study.**

Regime	Estimated Effect	χ^2_2	p-value
SSC [1]	-0.836 + 0.003×43	0.345	0.842
SSB [2]	-0.913 + 0.004×43		
SSM [3]	-0.711 + 0.0002×43		

[1] Treat with SERT, continue SERT if respond, otherwise add CBASP to SERT
[2] Treat with SERT, continue SERT if respond, otherwise add BSP to SERT
[3] Treat with SERT, continue SERT if respond, otherwise add MED to SERT

3.0 WEIGHTED ESTIMATING EQUATIONS IN TWO-STAGE LONGITUDINAL STUDIES IN THE PRESENCE OF TIME-DEPENDENT MISSING

3.1 INTRODUCTION

Pharmacotherapies and psychotherapies are commonly used to alleviate mood disorders in patients with chronic forms of major depression [Keller et al., 2000, Cuijpers et al., 2010]. The effect of using a single treatment in the management of depression had been investigated rigorously in medical research. The most common pharmacotherapies include classes of selective serotonin reuptake inhibitors (SSRIs), serotonin-norepinephrine reuptake inhibitors (SNRIs), tetracyclic antidepressants (TeCAs), and dopamine reuptake inhibitors (DRIs). The psychotherapies include cognitive behavioral analysis system of psychotherapy (CBASP) and brief supportive psychotherapy (BSP). Despite the apparent effectiveness of treatments for major depressive disorders, nearly 50% of the patients with chronic forms of major depressive disorders fail to respond to the first line of pharmacotherapies or psychotherapies [Kocsis et al., 2009]. Studies have shown that the augmentation with psychotherapy when pharmacotherapy alone was ineffective seemed to be an effective alternative to the treatment of chronically depressed patients [Cuijpers et al., 2009].

In an attempt to investigate the effect of sequence of pharmacotherapy and psychotherapy in the treatment of patients with chronic depression, the Research Evaluating the Value of Augmenting Medication with Psychotherapy (REVAMP) study [Trivedi et al., 2008] enrolled a total of 808 patients to determine the role of adjunctive psychotherapy in chronically depressed patients who had less than complete response to an initial medication. The study consisted of two stages of maximum 12-week duration, for a maximum treatment duration of

24 weeks. In the first stage, patients were assigned one of four treatments of antidepressants by the REVAMP physicians based on the algorithm using information on pharmacotherapy treatment history. These antidepressants were Sertaline (SERT), Escitalopram (EcCIT), Bupropion (BUP-SR), and Venlafaxine (VLF-XR). During weeks 8 through 12, patients whose depressive symptoms did not meet certain criteria were declared as non-responders. In the second stage, the non-responders were randomly assigned to three groups in a 2:2:1 ratio. The first two groups were to have Cognitive Behavioral Analysis System of Psychotherapy (CBASP) or Brief Supportive Psychotherapy (BSP) added to their pharmacotherapy and the third group was to receive medication alone (MED). Patients who were declared responders, continued receiving their antidepressant from the first stage. Throughout the study, a patient's 24-item Hamilton Rating Scale for Depression (HRSD) scores [Hamilton, 1960] were collected at each visit. The higher the HRSD scores, the worse is the symptoms of depression. The goal of the REVAMP study was to find the most beneficial combination of pharmacotherapy and psychotherapy for each chronically depressed patient.

In many depression studies, multiple stages of sequential treatments had been suggested as an alternative treatment strategy to a single episode of treatment for chronically depressive disorders [Nierenberg et al., 2003, Thase et al., 2007]. For example, a strategy for treating patients with chronically depressive disorders could be: treat with the antidepressant based on the pre-determined treatment algorithm in the first stage, continue the same if responds by 12 weeks, else add CBASP with the medication. This strategy in which the treatment in the second stage depends on a patient's intermediate response is called adaptive treatment strategy [Lavori and Dawson, 2000] or dynamic treatment regime [Murphy, 2003]. When physicians make decisions about treatments for patients with chronically depressive disorders, their decisions follow the logic of adaptive strategy or dynamic regime. However, most of the conventional statistical analysis focuses only on the comparisons among fixed treatments or non-dynamic regime. To estimate the effects of depression treatment regimes in two-stage longitudinal studies, adjustments to the conventional statistical analyses are needed [Lunceford et al., 2002].

Estimation of time-varying treatment effects has been discussed in the literature by many authors [Hernán et al., 2000, Murphy et al., 2001, Ko et al., 2003, Bodnar et al., 2004,

VanderWeele, 2009]. An estimation method for the mean response that would have been observed if the entire population followed a dynamic regime, where the available data were observational [Murphy et al., 2001]. A marginal structural model was used to estimate the causal effects of time-dependent treatments by modeling the counterfactual random variables [Hernán et al., 2000, Ko et al., 2003, Bodnar et al., 2004]. A marginal structural model with inverse probability treatment weighting was used to estimate the controlled direct and indirect effects [VanderWeele, 2009].

Missing data is common in almost all longitudinal studies. For example, in the two-stage longitudinal REVAMP study, missing data could occur in two ways. First, by study design, patients are randomized to three groups in the second stage, and hence patients receiving one psychotherapy will have missing (counterfactual) data on the other groups. Second, patients can drop out at any time during the study period. Under the framework of the generalized estimating equations [Liang and Zeger, 1986], the inverse-probability-weighted methodology could correct the bias caused by ignoring missing data [Robins and Rotnitsky, 1992, Robins et al., 1995]. With the knowledge of time to response and time to drop-out, one could also improve upon the analysis by including more observations in the analysis compared to complete case analysis, which may intuitively increase the efficiency of the estimators.

In this chapter, we demonstrate an application of the inverse-probability-weighted generalized estimating equations to a two-stage longitudinal depression study in order to estimate the effects of depression treatment regimes in the presence of drop-out. We constructed the weights which allow us to use the partial information from patients who belonged to treatment regimes other than the targeted one and who dropped out from the study. The outline of this paper is as follows. We introduce notation, model, and assumptions in Section 3.2. In Section 3.3, we show how to construct the weights and draw inference from the observed data. In Section 3.4, we evaluate the asymptotic properties of the proposed methods through simulations. In Section 3.5, we demonstrate these methods through an application to the REVAMP dataset. We wrap up with a discussion in Section 3.6. The figures and tables are listed in Section 3.7.

3.2 MODEL FRAMEWORK AND NOTATION

Let us consider a study with two-stage setting that mimics the design of the REVAMP study. For each patient i ($i = 1, \ldots, n$), there are J treatments available in the first stage, A_j, where $j \in \{1, \ldots, J\}$. Patients responding to A_j are assigned to treatment B_k, where $k \in \{1, \ldots, K\}$ and patients failing to respond to A_j are assigned to treatment B'_l, where $l \in \{1, \ldots, L\}$. In the REVAMP study, $J = 1$ (SERT), $K = 1$ (SERT), and $L = 3$ (CBASP, BSP, and MED). A treatment regime $A_j B_k B'_l$ is then defined as "treat with A_j followed by B_k if respond, by B'_l if otherwise." In longitudinal studies, patients are followed over time, and for patient i, let Y_{im} be the continuous outcome measured at time t_{im}, where $m \in \{1, \ldots, M_i\}$ and $\mathbf{Y}_i = \{Y_{i1}, \ldots, Y_{iM_i}\}^T$.

In the presence of randomization and drop-out, it is often useful to apply the idea of counterfactuals to the data analysis [Holland, 1986]. For patient i, define $R_i(A_j)$ to be the response status if he/she had received the first line treatment A_j; let $T_{im_{1i}}$ be the time when patient i is declared as a responder or a non-responder to the first line treatment A_j, at which point randomization to the second set of treatment occurs, $m_{1i} \in \{1, \ldots, M_i\}$. Whether observed or not, we define the following outcomes: $\mathbf{Y}_i(A_j)$, a $m_{1i} \times 1$ vector of repeated measures of outcome at time points in the first stage if patient i receives treatment A_j; $\mathbf{Y}_i(A_j B_k)$, a $(M_i - m_{1i}) \times 1$ vector of repeated measures of outcome at each time point in the second stage if patient i receives treatment A_j in the first stage and B_k in the second stage after responding to A_j; $\mathbf{Y}_i(A_j B'_l)$, a $(M_i - m_{1i}) \times 1$ vector of repeated measures of outcome at each time point in the second stage if patient i receives treatment A_j in the first stage and B'_l in the second stage after failing to respond to A_j. For simplicity, let us assume $J = K = L = 2$. Thus, for one initial treatment, A_1, patient i could be associated with the following random variables:

$$\{[R_i(A_1), T_{im_{1i}}], \mathbf{Y}_i(A_1), \mathbf{Y}_i(A_1 B_1), \mathbf{Y}_i(A_1 B_2), \mathbf{Y}_i(A_1 B'_1), \mathbf{Y}_i(A_1 B'_2)\}. \tag{3.1}$$

In terms of these counterfactual variables, we define $\mathbf{Y}_i(A_1 B_k B'_l)$ to be an $M_i \times 1$ vector outcome measurements under treatment regime $A_1 B_k B'_l$, for $k, l \in \{1, 2\}$. The m-th element

of $\mathbf{Y}_i(A_1B_kB_l')$ can be expressed as:

$$Y_{im}(A_1B_kB_l') = I(t_{im} \leq T_{im_{1i}})Y_{im}(A_1) \\ + I(t_{im} > T_{im_{1i}}) \left\{ R_i(A_1)Y_{i(m-m_{1i})}(A_1B_k) + [1 - R_i(A_1)]\, Y_{i(m-m_{1i})}(A_1B_l') \right\}. \quad (3.2)$$

The counterfactual outcome $\mathbf{Y}_i(A_2B_kB_l')$ can be defined in a similar fashion. Because patients who receive the initial treatment A_2 are independent samples to patients receiving A_1, without loss of generality, we will focus on one treatment A_1 in the first stage. Our interest is to estimate the effect of treatment regime $A_1B_kB_l'$ on the counterfactual outcome $Y_{im}(A_1B_kB_l')$ over time. In other words, we are focusing on the estimation of the coefficient $\beta_{1,1kl}$ in the following marginal mean model,

$$E[Y_{im}(A_1B_kB_l') \mid \mathbf{x}_{im}] = \mathbf{x}_{im}^T \boldsymbol{\beta}_{1kl}, \quad (3.3)$$

where $\mathbf{x}_{im}^T = [1, t_{im}, w_{1i}, \ldots, w_{pi}]$, $\{w_{1i}, \ldots, w_{pi}\} \in \mathbf{W}_i$, and $\boldsymbol{\beta}_{1kl}^T = [\beta_{0,1kl}, \ldots, \beta_{p+1,1kl}]$.

If $Y_{im}(A_1B_kB_l')$ was observed for each patient in the sample, any conventional statistical method could have provided valid estimation to these coefficients. However, in reality, we cannot observe the counterfactual outcome $Y_{im}(A_1B_kB_l')$ for all patients. For example, if a patient received A_1, responded to A_1, and then received B_2, we do not observe the counterfactual outcome $Y_{im}(A_1B_1B_l')$ for that patient. The complete observed data are characterized as the set of i.i.d. random list:

$$\{[R_i, T_{im_{1i}}], R_iZ_{ki}, (1 - R_i)Z_{li}', \mathbf{W}_i, \mathbf{Y}_i\}, \quad \text{for } k, l \in \{1, 2\},$$

where $R_i = 1$, if the patient is a responder, and $R_i = 0$ otherwise; $\mathbf{W}_i$ is a finite set of baseline covariates; Z_{ki} and Z_{li}' are the assignment indicators for treatment B_k and B_l', respectively, for $k, l \in \{1, 2\}$; $Z_{1i} = 1(0)$ if patient i is randomized to $B_1(B_2)$; $Z_{1i}' = 1(0)$ if patient i is randomized to $B_1'(B_2')$; Z_{2i} and Z_{2i}' satisfy $Z_{1i} + Z_{2i} = 1$ and $Z_{1i}' + Z_{2i}' = 1$; $\mathbf{Y}_i$ is a $M_i \times 1$ vector of repeated observed outcome for patient i.

In order to draw inference on $\mathbf{Y}_i(A_1B_kB_l')$ from the observed data $\mathbf{Y}_i$, one requires the consistency assumption (CA) to connect observed data and counterfactuals [Rubin, 1974, Robins et al., 2000]. The CA implies that the observed outcome is equal to the counterfactual

outcome under treatment assignment consistent with the counterfactual. In other words, the m-th element of $\mathbf{Y}_i$ is:

$$Y_{im} = I(t_{im} \leq T_{im_{1i}}) Y_{im}(A_1) + I(t_{im} > T_{im_{1i}}) \left\{ R_i \sum_{k=1}^{2} Z_{ki} Y_{i(m-m_{1i})}(A_1 B_k) + (1 - R_i) \sum_{l=1}^{2} Z'_{li} Y_{i(m-m_{1i})}(A_1 B'_l) \right\}. \quad (3.4)$$

Another frequently made assumption is the Sequential Randomization Assumption (SRA) which states that the probabilities of receiving treatment B_k and B'_l do not depend on counterfactuals given the history of information collected prior to the randomization [Rubin, 1974, Robins, 1986]:

$$P\{Z_{ki} = 1 \mid [R_i = 1, T_{im_{1i}}], \mathbf{W}_i, \mathbf{Y}_i(A_1 B_k B'_l)\} = P\{Z_{ki} = 1 \mid [R_i = 1, T_{im_{1i}}], \mathbf{W}_i\};$$
$$P\{Z'_{li} = 1 \mid [R_i = 0, T_{im_{1i}}], \mathbf{W}_i, \mathbf{Y}_i(A_1 B_k B'_l)\} = P\{Z'_{li} = 1 \mid [R_i = 0, T_{im_{1i}}], \mathbf{W}_i\};$$
$$\text{for } k, l \in \{1, 2\}. \quad (3.5)$$

In the REVAMP study, the second stage treatment assignment probabilities were constant. Therefore, we define $P\{Z_{ki} = 1 \mid [R_i = 1, T_{im_{1i}}], \mathbf{W}_i\} = \eta_k$ and $P\{Z'_{li} = 1 \mid [R_i = 0, T_{im_{1i}}], \mathbf{W}_i\} = \zeta_l$ to be constant as well.

3.3 INFERENCE

Our goal is to estimate the effects of treatment regimes $A_1 B_k B'_l$ where $k, l \in \{1, 2\}$ in reducing the HRSD scores under the generalized estimating equations (GEE) framework [Liang and Zeger, 1986]. If $\mathbf{Y}_i(A_1 B_k B'_l)$ was observed for all patients, a generalized estimating equation of the form

$$\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \{\mathbf{Y}_i(A_1 B_k B'_l) - \mathbf{X}_i \boldsymbol{\beta}_{1kl}\} = \mathbf{0}, \quad (3.6)$$

could have been used to estimate the effect of treatment regime $A_1 B_k B'_l$ on the HRSD scores [by the adjusted time coefficient (continuous time) or by the adjusted change (discrete time)], where $\mathbf{X}_i^T = [\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iM_i}]$, and $\mathbf{x}_{im}$ is defined as in (3.3); $\mathbf{V}_{i,1kl} = \phi \mathbf{A}_{i,1kl}^{1/2} \mathbf{R}(\boldsymbol{\alpha}) \mathbf{A}_{i,1kl}^{1/2}$; $\mathbf{R}(\boldsymbol{\alpha})$

is an $M_i \times M_i$ working correlation matrix specified by $\boldsymbol{\alpha}$; $\mathbf{A}_{i,1kl}$ is an $M_i \times M_i$ diagonal matrix with $v_{im}(\mathbf{X}_i; \boldsymbol{\beta}_{1kl})$ as the m-th element, where $v_{im}(\mathbf{X}_i, \boldsymbol{\beta}_{1kl})$ is the assumed working variance function of $Y_{im}(A_1 B_k B_l')$ and ϕ is the dispersion parameter. But in our setting, $\mathbf{Y}_i(A_1 B_k B_l')$ cannot be observed for all four possible treatment regimes $A_1 B_1 B_1'$, $A_1 B_1 B_2'$, $A_1 B_2 B_1'$, and $A_1 B_2 B_2'$. In this case, all patients belong to all four treatment regimes as long as they are in the first stage. In the second stage, each patient belongs to at most two treatment regimes but not all. For example, if a patient received A_1 in the first stage and received B_1 in the second stage after responding to A_1, he/she would be consistent with regimes $A_1 B_1 B_1'$ and $A_1 B_1 B_2'$ only. However, if a patient was randomized to A_1 at the first stage and dropped out before reaching the second stage, then he/she would be consistent with all four regimes starting with A_1. While estimating the effects of treatment regime $A_1 B_1 B_1'$, patients who receive B_2 or B_2' in the second stage would be considered to have missing data by study design (i.e. by randomization.) In the presence of missing data by randomization, the inverse-probability-weighting (IPW) methodology can provide valid estimation [Robins and Rotnitsky, 1992, Rotnitzky, 2009]. To account for missing data due to randomization in two-stage longitudinal studies, we construct the m-th element of the diagonal weight matrix as follows:

$$Q_{im,1kl} = I(t_{im} \leq T_{im_{1i}}) + I(t_{im} > T_{im_{1i}}) \times \left[\frac{R_i Z_{ki}}{\eta_k} + \frac{(1 - R_i) Z_{li}'}{\zeta_l} \right], \tag{3.7}$$

where $\eta_k = P(Z_{ki} = 1 \mid [R_i = 1, T_{im_{1i}}], \mathbf{W}_i)$, and $\zeta_l = P(Z_{li}' = 1 \mid [R_i = 0, T_{im_{1i}}], \mathbf{W}_i)$. The term $I(t_{im} \leq T_{im_{1i}})$ in (3.7) represents data from patient i before responding/failing to respond to A_1. By including the information of time to response, $T_{im_{1i}}$, all patients' information collected on or before time $T_{im_{1i}}$ will be used in the analysis, even if he/she was not consistent with the target treatment regime due to treatment assignment in the second stage. After the response status of patient i is confirmed, where $(t_{im} > T_{im_{1i}})$ in (3.7), his/her data will be inversely weighted by η_k or ζ_l based on the response status, R_i, and treatment allocation, Z_{ki} or Z_{li}', $[\eta_k^{-1} R_i Z_{ki} + \zeta_l^{-1}(1 - R_i) Z_{li}']$ in (3.7). Therefore, $Q_{im,1kl} \geq 1$ if patient i is consistent with regime $A_1 B_k B_l'$, otherwise $Q_{im,1kl} = 0$. The weight matrix $\mathbf{Q}_{i,1kl}$ is thus an $M_i \times M_i$ matrix, $\text{diag}(Q_{im,1kl}, m = 1, \ldots, M_i)$.

Lemma 3.3.1. *Under SRA in (3.5),* $E[\mathbf{Q}_{i,1kl} \mid \mathbf{X}_i, \mathbf{Y}_i(A_1 B_k B_l')] = \mathbf{I}_{Mi}$.

Proof. Using iterated conditioning,

$$
\begin{aligned}
&E\left[Q_{im,1kl} \mid \mathbf{X}_i, \mathbf{Y}_i(A_1 B_k B_l')\right] \\
&= E\left\{ I(t_{im} \leq T_{im_{1i}}) + I(t_{im} > T_{im_{1i}}) \left[\frac{R_i Z_{ki}}{\eta_k} + \frac{(1-R_i) Z_{li}'}{\zeta_l} \right] \mid \mathbf{X}_i, \mathbf{Y}_i(A_1 B_k B_l') \right\} \\
&= E\left(E\left\{ I(t_{im} \leq T_{im_{1i}}) + I(t_{im} > T_{im_{1i}}) \left[\frac{R_i Z_{ki}}{\eta_k} + \frac{(1-R_i) Z_{li}'}{\zeta_l} \right] \mid [R_i, T_{im_{1i}}], \mathbf{X}_i, \mathbf{Y}_i(A_1 B_k B_l') \right\} \right) \\
&= E\left\{ I(t_{im} \leq T_{im_{1i}}) + I(t_{im} > T_{im_{1i}}) \times E\left[\frac{R_i Z_{ki}}{\eta_k} + \frac{(1-R_i) Z_{li}'}{\zeta_l} \mid [R_i, T_{im_{1i}}], \mathbf{X}_i, \mathbf{Y}_i(A_1 B_k B_l') \right] \right\} \\
&= E\left[I(t_{im} \leq T_{im_{1i}}) + I(t_{im} > T_{im_{1i}}) \right] \qquad \text{(by SRA in (3.5))} \\
&= 1.
\end{aligned}
$$

□

In longitudinal studies, drop-out is a common phenomenon. To account for the partial information observed from patients who dropped out from the study, we re-define the m-th element of the diagonal weight matrix from (3.7) as

$$
Q_{im,1kl}(\overline{\pi}_{im}) = \frac{\Delta_{im}}{\overline{\pi}_{im}} \times \left\{ I(t_{im} \leq T_{im_{1i}}) + I(t_{im} > T_{im_{1i}}) \times \left[\frac{R_i Z_{ki}}{\eta_k} + \frac{(1-R_i) Z_{li}'}{\zeta_l} \right] \right\},
$$

where Δ_{im} is the indicator for observing outcome data from patient i at time t_{im}, i.e. $\Delta_{im} = 1$ if observed and $\Delta_{im} = 0$ if otherwise. Let π_{im} denote the probability of observing the outcome for patient i at time t_{im} given that the patient had data up to time $t_{i(m-1)}$, where $\pi_{im} = P(\Delta_{im} = 1 \mid \Delta_{i(m-1)} = 1)$, $m > 1$. Thus, $\overline{\pi}_{im}$ is the unconditional probability of observing the outcome at the m-th time point for the i-th patient, where $\overline{\pi}_{im} = \prod_{s=2}^{m} \pi_{is}$. Note that we assume monotone missing, i.e. $\Delta_{im} = 0$ implies $\Delta_{i(m+1)} = \ldots = \Delta_{iM_i} = 0$. We also assume that all outcome data are observed at baseline, i.e. $\Delta_{i1} = 1$ and $\pi_{i1} = 1$ for all patients. Therefore, each patient who was consistent with the treatment regime and had observed data at time t_{im} is additionally and inversely weighted by $\overline{\pi}_{im}$.

Lemma 3.3.2. *Under SRA in (3.5), and when π_{im} is known,* $E[\mathbf{Q}_{i,1kl}(\overline{\pi}_{im}) \mid \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB_l')] = \mathbf{I}_{Mi}$.

Proof. Using iterated conditioning,

$$
\begin{aligned}
& E\left[Q_{im,1kl}(\overline{\pi}_{im}) \mid \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB_l')\right] \\
&= E\left\{E\left[\frac{\Delta_{im}}{\overline{\pi}_{im}}Q_{im,1kl} \mid \Delta_{im}, [R_i, T_{im_{1i}}], \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB_l')\right]\right\} \\
&= E\left\{\frac{\Delta_{im}}{\overline{\pi}_{im}}E\left[Q_{im,1kl} \mid \Delta_{im}, [R_i, T_{im_{1i}}], \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB_l')\right]\right\} \\
&= E\left[\frac{\Delta_{im}}{\overline{\pi}_{im}}\right] \quad \text{(by Lemma 3.3.1)} \\
&= E\left\{\frac{1}{\overline{\pi}_{im}}E\left[\Delta_{im} | \mathbf{X}_i, \mathbf{Y}_i(A_1B_kB_l')\right]\right\} \\
&= \frac{1}{\overline{\pi}_{im}} \times \prod_{s=2}^{m}\pi_{is} \\
&= 1, \text{ for } m = 2, \ldots, M_i. \qquad \square
\end{aligned}
$$

In most of the cases, π_{im} is unknown and needs to be estimated. Let $\pi_{im}(\boldsymbol{\gamma}) = G_{im}(\boldsymbol{\gamma})$ be a postulated model for drop-out process defined by a set of parameters $\boldsymbol{\gamma}$. Let $\hat{\boldsymbol{\gamma}}$ be a regular and asymptotically linear (RAL) estimator of $\boldsymbol{\gamma}$. For each patient i who is consistent with treatment regime $A_1B_kB_l'$, we define the weight matrix, $\mathbf{Q}_{i,1kl}(\hat{\boldsymbol{\gamma}})$, as the $M_i \times M_i$ diagonal matrix, where each m-th diagonal element is $Q_{im,1kl}(\hat{\boldsymbol{\gamma}}) = \overline{\pi}_{im}^{-1}(\hat{\boldsymbol{\gamma}})\Delta_{im}Q_{im,1kl}$, where $\overline{\pi}_{im}(\hat{\boldsymbol{\gamma}}) = \prod_{s=2}^{m} G_{im}(\hat{\boldsymbol{\gamma}})$. This modified weight is then used in the estimating equation (3.6). Specifically, the weighted estimating equations can be expressed as

$$
\sum_{i=1}^{n}\mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \hat{\boldsymbol{\gamma}}\right) = \sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}(\hat{\boldsymbol{\gamma}})\left\{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right\} = \mathbf{0}. \tag{3.8}
$$

The solution of (3.8) can be obtained through the following iterative algorithm [Liang and Zeger, 1986]:

$$
\hat{\boldsymbol{\beta}}_{1kl}^{(r+1)} = \hat{\boldsymbol{\beta}}_{1kl}^{(r)} + \left[\sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}(\hat{\boldsymbol{\gamma}})\mathbf{X}_i\right]^{-1}\sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}(\hat{\boldsymbol{\gamma}})\left[\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_{1kl}^{(r)}\right]. \tag{3.9}
$$

Lemma 3.3.3. *Under CA in (3.4), SRA in (3.5), and when $\pi_{im} = G_{im}(\boldsymbol{\gamma})$ is correctly specified, $E\left[\mathbf{U}_i(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma})\right] = \mathbf{0}$.*

Proof. Under CA in (3.4), $\mathbf{Q}_{i,1kl}(\boldsymbol{\gamma})\mathbf{Y}_i = \mathbf{Q}_{i,1kl}(\boldsymbol{\gamma})\mathbf{Y}_i(A_1 B_k B_l')$. Thus,

$$
\begin{aligned}
E\left[\mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right)\right] &= E\left[\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \mathbf{Q}_{i,1kl}(\boldsymbol{\gamma}) \left\{\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}_{1kl}\right\}\right] \\
&= E\left[\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \mathbf{Q}_{i,1kl}(\boldsymbol{\gamma}) \left\{\mathbf{Y}_i(A_1 B_k B_l') - \mathbf{X}_i \boldsymbol{\beta}_{1kl}\right\}\right] \quad \text{(by CA in (3.4))} \\
&= E\left\{E\left[\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \mathbf{Q}_{i,1kl}(\boldsymbol{\gamma}) \left\{\mathbf{Y}_i(A_1 B_k B_l') - \mathbf{X}_i \boldsymbol{\beta}_{1kl}\right\} \middle| \mathbf{X}_i, \mathbf{Y}_i(A_1 B_k B_l')\right]\right\} \\
&= E\left[\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \left\{\mathbf{Y}_i(A_1 B_k B_l') - \mathbf{X}_i \boldsymbol{\beta}_{1kl}\right\}\right] \quad \text{(by Lemma 3.3.2)} \\
&= E\left\{E\left[\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \left\{\mathbf{Y}_i(A_1 B_k B_l') - \mathbf{X}_i \boldsymbol{\beta}_{1kl}\right\} | \mathbf{X}_i\right]\right\} \\
&= E\left\{\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \left\{E\left[\mathbf{Y}_i(A_1 B_k B_l') | \mathbf{X}_i\right] - \mathbf{X}_i \boldsymbol{\beta}_{1kl}\right\}\right\} = \mathbf{0} \quad \text{(by Equation (3.3))}.
\end{aligned}
$$

□

Proposition 3.3.1. *$\hat{\boldsymbol{\beta}}_{1kl}$ is a consistent estimator of $\boldsymbol{\beta}_{1kl}$, provided (1) $\hat{\boldsymbol{\gamma}}$ is $\sqrt{n}$-consistent, (2) $\pi_{im} = G_{im}(\boldsymbol{\gamma})$ is specified correctly, and (3) π_{im} does not depend on the counterfactuals.*

Proof. From (1), we can write $\sqrt{n}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \boldsymbol{\varphi}_i(\boldsymbol{\gamma}) + \mathbf{o_p}(\mathbf{1})$, where $\boldsymbol{\varphi}_i(\boldsymbol{\gamma})$ is the influence function of the estimator $\hat{\boldsymbol{\gamma}}$ and $E[\boldsymbol{\varphi}_i(\boldsymbol{\gamma})] = \mathbf{0}$. Since $\hat{\boldsymbol{\beta}}_{1kl}$ satisfies Equation (3.8), the estimator $(\hat{\boldsymbol{\beta}}_{1kl}, \hat{\boldsymbol{\gamma}})^T$ is an M-estimator [Stefanski and Boos, 2002] defined by

$$
\sum_{i=1}^{n} \boldsymbol{\Psi}_i(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \left\{ \begin{array}{c} \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right) \\ \boldsymbol{\varphi}_i(\boldsymbol{\gamma}) \end{array} \right\} = \mathbf{0}.
$$

The consistency of $\hat{\boldsymbol{\beta}}_{1kl}$ follows from the fact that $E[\boldsymbol{\Psi}_i(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma})] = \mathbf{0}$ (by Lemma 3.3.3). □

Proposition 3.3.2. *Under assumptions similar to Proposition 3.3.1, $\hat{\boldsymbol{\beta}}_{1kl}$ is an asymptotically normally distributed with mean $\boldsymbol{\beta}_{1kl}$ and variance Σ/n, where*

$$
\boldsymbol{\Sigma} = \left[E\left(\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \mathbf{X}_i\right)\right]^{-1} \mathbf{B}\left(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right) \left[E\left(\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \mathbf{X}_i\right)\right]^{-1}, \text{ and} \tag{3.10}
$$

$$
\begin{aligned}
\mathbf{B}\left(\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right) =& E\left[\mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right) \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right)^T\right] \\
&+ E\left[\mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right) \boldsymbol{\varphi}_i(\boldsymbol{\gamma})^T\right] \times E\left[\frac{\partial}{\partial \gamma^T} \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right)\right]^T \\
&+ E\left[\frac{\partial}{\partial \gamma^T} \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right)\right] \times E\left[\boldsymbol{\varphi}_i(\boldsymbol{\gamma}) \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right)^T\right] \\
&+ E\left[\frac{\partial}{\partial \gamma^T} \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right)\right] \times E\left[\boldsymbol{\varphi}_i(\boldsymbol{\gamma}) \boldsymbol{\varphi}_i(\boldsymbol{\gamma})^T\right] \times E\left[\frac{\partial}{\partial \gamma^T} \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right)\right]^T .
\end{aligned}
$$

Proof. We start with the fact that the estimator $\hat{\boldsymbol{\beta}}_{1kl}$ satisfies $\sum_{i=1}^n \mathbf{U}_i(\hat{\boldsymbol{\beta}}_{1kl}, \hat{\boldsymbol{\gamma}}) = \mathbf{0}$. First, expanding $\sum_{i=1}^n \mathbf{U}_i(\hat{\boldsymbol{\beta}}_{1kl}, \hat{\boldsymbol{\gamma}})$ around $\boldsymbol{\beta}_{1kl}$ using Taylor's expansion, we obtain

$$
n^{\frac{1}{2}}\left(\hat{\boldsymbol{\beta}}_{1kl} - \boldsymbol{\beta}_{1kl}\right) = \left[-\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}_{1kl}^T} \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \hat{\boldsymbol{\gamma}}\right)\right]^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \hat{\boldsymbol{\gamma}}\right) + \mathbf{o_p}(\mathbf{1}) \tag{3.11}
$$

Then, applying Taylor's expansion on (3.11) around $\boldsymbol{\gamma}$, (3.11) can be rewritten as

$$
\begin{aligned}
n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_{1kl} - \boldsymbol{\beta}_{1kl}) =& \left[-\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}_{1kl}^T} \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right)\right]^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right) \\
&+ \left\{\left[-\frac{1}{n}\sum_{i=1}^n \frac{\partial^2}{\partial \boldsymbol{\beta}_{1kl}^T \partial \boldsymbol{\gamma}^T} \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right)\right]^{-1} \frac{1}{n}\sum_{i=1}^n \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right)\right. \\
&\left. + \left[-\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}_{1kl}^T} \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right)\right]^{-1} \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\gamma}^T} \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right)\right\} \times n^{\frac{1}{2}}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) + \mathbf{o_p}(\mathbf{1}).
\end{aligned} \tag{3.12}
$$

As $n \to \infty$,

$$
-\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}_{1kl}^T} \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right) \to E\left(\mathbf{X}_i^T \mathbf{V}_{i,1kl}^{-1} \mathbf{X}_i\right) \tag{3.13}
$$

$$
\frac{1}{n}\sum_{i=1}^n \mathbf{U}_i\left(\beta_{1kl}, \gamma\right) \to \mathbf{0} \quad \text{(by Lemma 3.3.3), and} \tag{3.14}
$$

$$
\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \gamma^T} \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right) \to E\left[\frac{\partial}{\partial \gamma^T} \mathbf{U}_i\left(\boldsymbol{\beta}_{1kl}, \boldsymbol{\gamma}\right)\right]. \tag{3.15}
$$

Since $\hat{\boldsymbol{\gamma}}$ is $\sqrt{n}$-consistent, we can write $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) = n^{-1/2}\sum_{i=1}^n \boldsymbol{\varphi}_i(\boldsymbol{\gamma}) + \mathbf{o_p}(\mathbf{1})$, where $\boldsymbol{\varphi}_i(\boldsymbol{\gamma})$ is the influence function of the estimator $\hat{\boldsymbol{\gamma}}$ and $E[\boldsymbol{\varphi}_i(\boldsymbol{\gamma})] = \mathbf{0}$. Using (3.13)–(3.15), we can

rewrite (3.12) as $n^{1/2}(\hat{\boldsymbol{\beta}}_{1kl} - \boldsymbol{\beta}_{1kl}) = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\psi}_{i,1kl} + \mathbf{o_p}(\mathbf{1})$ where $\boldsymbol{\psi}_{i,1kl}$ is the influential function of $\hat{\boldsymbol{\beta}}_{1kl}$ and it can be expressed as

$$\boldsymbol{\psi}_{i,1kl} = \left[E\left(\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{X}_i\right)\right]^{-1}\left\{\mathbf{U}_i\left(\boldsymbol{\beta}_{1kl},\boldsymbol{\gamma}\right) + E\left[\frac{\partial}{\partial\gamma^T}\mathbf{U}_i\left(\boldsymbol{\beta}_{1kl},\boldsymbol{\gamma}\right)\right]\boldsymbol{\varphi}_i(\boldsymbol{\gamma})\right\}. \tag{3.16}$$

By the central limit theorem, we can deduce that $n^{1/2}(\hat{\boldsymbol{\beta}}_{1kl} - \boldsymbol{\beta}_{1kl}) \xrightarrow{d} \text{MVN}\left(\mathbf{0},\boldsymbol{\Sigma}\right)$ where $\boldsymbol{\Sigma} = E(\boldsymbol{\psi}_{i,1kl}\boldsymbol{\psi}_{i,1kl}^T)$ and is given in (3.10).

Let $\boldsymbol{\gamma}$ be a $(q+1)\times 1$ vector, i.e. $\boldsymbol{\gamma} = [\gamma_0,\ldots,\gamma_q]^T$; therefore, $E[\frac{\partial}{\partial\gamma^T}\mathbf{U}_i\left(\boldsymbol{\beta}_{1kl},\boldsymbol{\gamma}\right)]$ will be a $(q+1)\times(p+2)$ matrix. Each column of matrix $E[\frac{\partial}{\partial\gamma^T}\mathbf{U}_i\left(\boldsymbol{\beta}_{1kl},\boldsymbol{\gamma}\right)]$ is the result of the partial derivative with respect to $\gamma_0,\ldots,\gamma_q$, respectively. Therefore,

$$\begin{aligned}
&E\left[\frac{\partial}{\partial\gamma^T}\mathbf{U}_i\left(\boldsymbol{\beta}_{1kl},\boldsymbol{\gamma}\right)\right]\\
&= E\left[\frac{\partial}{\partial\gamma_0}\mathbf{U}_i\left(\boldsymbol{\beta}_{1kl},\boldsymbol{\gamma}\right);\ldots;\frac{\partial}{\partial\gamma_q}\mathbf{U}_i\left(\boldsymbol{\beta}_{1kl},\boldsymbol{\gamma}\right)\right]\\
&= E\left[\frac{\partial}{\partial\gamma_0}\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}(\boldsymbol{\gamma})\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right);\ldots;\frac{\partial}{\partial\gamma_q}\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\mathbf{Q}_{i,1kl}(\boldsymbol{\gamma})\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right)\right]\\
&= E\left\{\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\left[\frac{\partial}{\partial\gamma_0}\mathbf{Q}_{i,1kl}(\boldsymbol{\gamma})\right]\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right);\ldots;\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\left[\frac{\partial}{\partial\gamma_q}\mathbf{Q}_{i,1kl}(\boldsymbol{\gamma})\right]\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right)\right\}\\
&= E\left\{\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\boldsymbol{\Gamma}_{i0}(\boldsymbol{\gamma})\mathbf{Q}_{i,1kl}(\boldsymbol{\gamma})\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right);\ldots;\mathbf{X}_i^T\mathbf{V}_{i,1kl}^{-1}\boldsymbol{\Gamma}_{iq}(\boldsymbol{\gamma})\mathbf{Q}_{i,1kl}(\boldsymbol{\gamma})\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_{1kl}\right)\right\},
\end{aligned}$$

where

$$\boldsymbol{\Gamma}_{is}(\boldsymbol{\gamma}) = \text{diag}\left\{\left[-\sum_{v=1}^{1}G_{iv}^{-1}(\boldsymbol{\gamma})\frac{\partial}{\partial\gamma_s}G_{iv}(\boldsymbol{\gamma})\right],\ldots,\left[-\sum_{v=1}^{M_i}G_{iv}^{-1}(\boldsymbol{\gamma})\frac{\partial}{\partial\gamma_s}G_{iv}(\boldsymbol{\gamma})\right]\right\},$$

for $s = 0,\ldots,q$. □

3.4 SIMULATION STUDY

To evaluate the performance of our proposed IPWGEE estimators in small samples, we conducted several simulation studies. We simulated data from a population which has similar design to the REVAMP study. Each patient had repeated measurements at seven visits such that $\mathbf{t}_i^T = [t_{i1}, \ldots, t_{i7}] = [0, 2, 4, 6, 8, 10, 12]$ for patient i. We considered one initial treatment A_1, two second stage treatments, B_1 and B_2, for responders, and two second stage treatments, B_1' and B_2', for non-responders. All patients received initial treatment A_1 at $t_{i1} = 0$. For each patient i, we generated counterfactual random variables listed in (3.1). Similar to the REVAMP study, patients were assessed to see if they had responded to the initial treatment A_1 at $t_{i3} = 4$ and $t_{i4} = 6$. Thus, the index of the time of response in this case is $m_{1i} = 3$ or 4. The response status, R_i, was drawn from a Bernoulli(θ) distribution. We assume that of these responders, $100 \times \theta_1\%$ responded at time t_{i3} and the rest at time t_{i4}. Thus each responder was assigned randomly to respond at time t_{i3} with probability θ_1. We considered two sets of parameters, $(\theta, \theta_1) = (0.5, 0.25)$ and $(0.3, 0.15)$ meaning 50% (30%) of the subjects would respond of whom 25% (15%) would respond at time t_{i3} and the rest at time t_{i4}. We also generated a baseline covariate, age, which follows a normal distribution with mean $\mu_{age} = 45$ and standard deviation $\sigma_{age} = 11$. For each patient i, counterfactual outcome vectors $\mathbf{Y}_i(A_1)$, $\mathbf{Y}_i(A_1B_1)$, $\mathbf{Y}_i(A_1B_2)$, $\mathbf{Y}_i(A_2B_1)$, and $\mathbf{Y}_i(A_2B_2')$, were generated from the following multivariate normal distributions (MVN):

$$
\begin{aligned}
\mathbf{Y}_i(A_1) &\sim \text{MVN}_{m_{1i}}((\theta_{1,A_1} + \theta_{2,A_1} \times \text{age}_i) \times \mathbf{1}_{m_{1i}} + \theta_{3,A_1} \times \mathbf{t}_{0i}, \mathbf{\Sigma}_0), \\
\mathbf{Y}_i(A_1B_1) &\sim \text{MVN}_{7-m_{1i}}((\theta_{1,A_1B_1} + \theta_{2,A_1B_1} \times \text{age}_i) \times \mathbf{1}_{(7-m_{1i})} + \theta_{3,A_1B_1} \times \mathbf{t}_{1i}, \mathbf{\Sigma}_1), \\
\mathbf{Y}_i(A_1B_2) &\sim \text{MVN}_{7-m_{1i}}((\theta_{1,A_1B_2} + \theta_{2,A_1B_2} \times \text{age}_i) \times \mathbf{1}_{(7-m_{1i})} + \theta_{3,A_1B_2} \times \mathbf{t}_{1i}, \mathbf{\Sigma}_1), \\
\mathbf{Y}_i(A_1B_1') &\sim \text{MVN}_{7-m_{1i}}((\theta_{1,A_1B_1'} + \theta_{2,A_1B_1'} \times \text{age}_i) \times \mathbf{1}_{(7-m_{1i})} + \theta_{3,A_1B_1'} \times \mathbf{t}_{1i}, \mathbf{\Sigma}_1), \\
\text{and } \mathbf{Y}_i(A_1B_2') &\sim \text{MVN}_{7-m_{1i}}((\theta_{1,A_1B_2'} + \theta_{2,A_1B_2'} \times \text{age}_i) \times \mathbf{1}_{(7-m_{1i})} + \theta_{3,A_1B_2'} \times \mathbf{t}_{1i}, \mathbf{\Sigma}_1),
\end{aligned}
$$

where $\mathbf{t}_{0i}$ is an $m_{1i} \times 1$ vector which indicates visits of outcome measurements in the first stage and $\mathbf{t}_{1i}$ is a $(7 - m_{1i}) \times 1$ vector which indicates visits of outcome measurements in the second stage; $\mathbf{\Sigma}_0(\sigma_0, \rho_0)$ is an $m_{1i} \times m_{1i}$ covariance matrix defined by standard deviation

σ_0 and autoregressive correlation ρ_0; $\boldsymbol{\Sigma}_1(\sigma_1, \rho_1)$ is a $(7 - m_{1i}) \times (7 - m_{1i})$ covariance matrix defined by standard deviation σ_1 and autoregressive correlation ρ_1. The values of σ_0, σ_1, ρ_0, and ρ_1 were set to be 5, 3, 0.8, and 0.8, respectively. For each patient i, the m-th element of the counterfactual outcome under a treatment regime $A_1 B_k B_l'$ for $k, l \in \{1, 2\}$ is generated by (3.2). The main focus in this study was to estimate the parameters $\boldsymbol{\beta}_{1kl}$ in the model $E[Y_{im}(A_1 B_k B_l') \mid \mathbf{x}_{im}] = \mathbf{x}_{im}^T \boldsymbol{\beta}_{1kl}$, where $\mathbf{x}_{im} = [1, t_{im}, \text{age}_i]^T$.

We considered the following parameter values for each counterfactual vector: θ_{1,A_1}= 25, θ_{2,A_1}= 0.5, θ_{3,A_1}= -0.5, θ_{1,A_1B_1}= 27, θ_{2,A_1B_1}= 0.6, θ_{3,A_1B_1}= -1.5, θ_{1,A_1B_2}= 38, θ_{2,A_1B_2}= 0.4, θ_{3,A_1B_2}= -2, $\theta_{1,A_1B_1'}$= 36, $\theta_{2,A_1B_1'}$= 0.7, $\theta_{3,A_1B_1'}$= -3, $\theta_{1,A_1B_2'}$= 68, $\theta_{2,A_1B_2'}$= 0.3, and $\theta_{3,A_1B_2'}$= -5. For this population, the true parameter values were $\boldsymbol{\beta}_{111}^T = [23.09, -1.13, 0.57]$, $\boldsymbol{\beta}_{112}^T = [27.94, -1.53, 0.48]$, $\boldsymbol{\beta}_{121}^T = [25.45, -1.30, 0.52]$, and $\boldsymbol{\beta}_{122}^T = [30.38, -1.69, 0.43]$ for 50% response, and $\boldsymbol{\beta}_{111}^T = [22.88, -1.20, 0.58]$, $\boldsymbol{\beta}_{112}^T = [29.65, -1.77, 0.45]$, $\boldsymbol{\beta}_{121}^T = [24.28, -1.31, 0.55]$, and $\boldsymbol{\beta}_{122}^T = [31.10, -1.86, 0.42]$ for 30% response.

2000 Monte Carlo samples of sizes 250, 500, and 900 observations were drawn from the populations described above with the following characteristics. For responders (i.e. $R_i = 1$), the assignment indicator for treatment B_1, Z_{1i}, was drawn from a Bernoulli distribution with probability η_1; the assignment indicator for treatment B_2 was defined as $Z_{2i} = 1 - Z_{1i}$. For non-responders (i.e. $R_i = 0$), the assignment indicator for treatment B_1', Z_{1i}' was drawn from a Bernoulli distribution with probability ζ_1; the assignment indicator for treatment B_2' was defined as $Z_{2i}' = 1 - Z_{1i}'$. We assumed parameters $\{\eta_1, \zeta_1\}$ to be $\{0.5, 0.5\}$. Then, we defined the observed outcome for patient i at time t_{im} as in (3.4).

Additionally, a number of patients were allowed to drop out at each time point based on the following logistic regression model: $\pi_{im} = G_{im}(\boldsymbol{\gamma}) = \{1 + \exp[-\gamma_0 - \gamma_1 \times t_{im} - \gamma_2 \times \text{age}_i - \gamma_3 \times Y_{i(m-1)} - \gamma_4 \times t_{im} \times \text{age}_i - \gamma_5 \times t_{im} \times Y_{i(m-1)}]\}^{-1}$. Choices of parameters $\boldsymbol{\gamma}^T = [2.1, 0.11, 0.01, -0.02, 0.001, -0.002]$ and $\boldsymbol{\gamma}^T = [2.42, 0.1, 0.02, -0.04, 0.002, -0.006]$ gave us approximate drop-out rates of 30% and 50%, respectively.

Table 3.1 shows the simulation results of the IPWGEE estimators under the scenarios of 50% response rate. When the drop-out rate was 30%, the IPWGEE estimators were approximately unbiased for all regimes with a sample size of 250. The maximum relative bias observed was 0.8%. The estimated standard errors of the estimators were consistent with

the Monte Carlo standard errors. The coverage probabilities for the 95% Wald confidence intervals for the parameters were ranged between 94.1% and 94.9%. When the sample size was increased, the results remained the same for all regimes as expected. When the drop-out rate was 50% and the sample size was 250, the maximum relative bias of the estimators increased to 1.8%. The estimated standard errors of the estimators were slightly inconsistent with the Monte Carlo standard deivations. When the sample size was increased to 500 and 900, the estimators were approximately unbiased for all regimes with the maximum relative bias of 1.3%. The estimated standard errors of the estimators were consistent with the Monte Carlo standard errors. For the sample size of 900, the coverage probabilities for the 95% Wald confidence intervals for the parameters were ranged between 91.9% and 94.3%.

For the scenarios of 30% response rate (shown in Table 3.2), the IPWGEE estimators were approximately unbiased for all regimes regardless the size of the sample. The maximum relative biases observed were 1.5% for 30% drop-out rate and 1.7% for 50% drop-out rate. Similar to the resutls presented in Table 3.1, the estimated standard errors of the estimators were consistent with the Monte Carlo standard errors when the drop-out rate was 30%. For the drop-out rate of 50%, a large size of sample was needed in order to observe the consistency between the estimated standard errors of the estimators and the Monte Carlo standard errors. For the sample size of 900, the coverage probabilities for the 95% Wald confidence intervals for the parameters were ranged between 91.9% and 93.3% for 30% drop-out and ranged between 91.7% and 94.3% for 50% drop-out.

In both Tables 3.1 and 3.2, the IPWGEE estimators gained efficiency with increasing sample size. The estimated standard errors of the estimators and the Monte Carlo standard errors were closer to each other when the sample size became larger. When the drop-out rate increased, the IPWGEE estimators required a larger sample size in order to achieve the consistency.

Table 3.3 shows the comparisons of the IPWGEE estimators based on subject-specific and time-dependent weights under the scenarios of 50% response rate. Regardless the drop-out rates, both weighting methods provided unbiased estimators among four treatment regimes. Because of the additional information incorporated in the time-dependent weighting method, its Monte Carlo standard errors were relatively smaller than the Monte Carlo standard errors

provided by subject-specific weighting method. For example, for regime $A_1B_1B_1'$ with a sample size of 250 and 30% drop-out rate, the Monte Carlo standard error was 0.070 for subject-specific weighting method and 0.052 for time-dependent weighting method. Similar results were found under the scenarios of 30% response rate (Table 3.4).

3.5 ANALYSIS OF THE REVAMP STUDY

A total of 618 patients with chronic forms of major depressive disorders received Sertraline (SERT) as the initial treatment in the REVAMP study for a maximum of 12 weeks [Trivedi et al., 2008]. During weeks 8 through 12 in the first stage, the response status to SERT from each patient was determined. Patients responding to SERT continued receiving SERT in the second stage for another 12 weeks. Patients failing to respond to SERT were randomized to additionally receive one the the following three treatments for another 12 weeks: (1) Cognitive Behavior Analysis System of Psychotherapy (CBASP), (2) Brief Supportive Psychotherapy (BSP), or (3) Medication alone (MED). The randomization was done with probabilities of 0.4, 0.4, and 0.2, respectively, to CBASP, BSP, and MED. The outcome of the study was measured by the 24-item Hamilton Rating Scale for Depression (HRSD) scores at each visit. A reduction of the HRSD scores from baseline would indicate that the patient was recovering from depression.

Figure 3.1 depicts the design and patient flow of 618 patients in the REVAMP study. In the first stage, 125 patients responded to SERT and 357 patients failed to respond to SERT. In the first stage, 136 patients dropped out from the study at different points of time prior to ascertaining their response status. In the second stage, 125 responders continued receiving SERT and 101 of them had; thus 24 patients' HRSD scores were missing at one or more time points. Among 357 patients who did not respond to SERT, 137 received CBASP (120 of them had complete HRSD scores), 147 received BSP (123 of them had complete HRSD scores), and 73 received MED (56 of them had complete HRSD scores). Overall, 218 patients (136 in the first stage and 82 in the second stage) did not have complete HRSD scores but had partial observations prior to drop-out. Figure 3.2 illustrates how the HRSD

scores changed over time in different groups. For responders (Figure 3.2(a)), their HRSD scores dropped rapidly from the beginning to the end of study. For non-responders, the reductions of the HRSD scores from baseline were similar to each other among CBASP, BSP, and MED (Figures 3.2(b), 3.2(c), and 3.2(d), respectively), and their slopes were much smaller compared to responders' slope. For drop-outs (Figure 3.2(e)), the reduction of the HRSD scores from baseline showed a flat trend and the variation was the largest among five groups.

Our goal was to estimate the effects of all possible depression treatment regimes in reducing the HRSD scores in the REVAMP study. Hence, starting with initial treatment SERT, there are three depression treatment regimes for which the REVAMP study: (1) SSC: treat with SERT, continue SERT if respond, otherwise add CBASP to SERT; (2) SSB: treat with SERT, continue SERT if respond, otherwise add BSP to SERT; and (3) SSM: treat with SERT, continue SERT if respond, otherwise add MED to SERT. Therefore, responders with SERT in Figure 3.2(a) and non-responders with CBASP in Figure 3.2(b) are consistent with depression treatment regime SSC. Responders with SERT in Figure 3.2(a) and non-responders with BSP in Figure 3.2(c) are consistent with depression treatment regime SSB. Responders with SERT in Figure 3.2(a) and non-responders with MED in Figure 3.2(d) are consistent with depression treatment regime SSM. The effects of depression treatment regime r, where $r \in \{SSC, SSB, SSM\}$, could be formulated as the coefficient $\boldsymbol{\beta}_r$ in the marginal mean model $E[\text{HRSD}_{im}(r) \mid t_{im}, w_{1i}, \ldots, w_{pi}] = \beta_{0,r} + \beta_{1,r} \times t_{im} + \beta_{2,r} \times w_{1i} + \ldots + \beta_{p+1,r} \times w_{pi}$, where $\{w_{1i}, \ldots, w_{pi}\} \in \mathbf{W}_i$. The selection of baseline covariates in the marginal mean model were determined by the analysts through knowledge about the study or the standard model-building process. We chose age and employment status as the baseline covariates along with an interaction between time and age in the marginal mean model. Hence, the marginal mean model for depression treatment regime r is

$$E\left[\text{HRSD}_{im}(r) \mid t_{im}, \text{age}_i, \text{emp}_i\right] = \beta_{0,r} + \beta_{1,r} \times t_{im} + \beta_{2,r} \times \text{age}_i + \beta_{3,r} \times \text{emp}_i + \beta_{4,r} \times t_{im} \times \text{age}_i,$$

where emp_i=1 if employed and emp_i=0 if otherwise. The inference was drawn from the observed outcome Y_{im} using the IPWGEE methodology.

Figure 2(e) shows data from 218 patients who had dropped out from the REVAMP study. Their HRSD scores were not available starting after drop-out. However, these patients' HRSD scores collected prior to drop-out could provide useful partial information. Table 3.5 shows the distributions of drop-outs across baseline covariates. In univariate analysis using chi-square test (or Fisher's exact test) with level of 0.05, Caucasian race, higher education, and older age were significantly associated with having complete data. To account for the incompleteness through the IPWGEE methodology, the conditional probability of observing the HRSD scores for patient i at time t_{im}, π_{im}, was estimated from the sample based on a logistic regression model. Based on the results in Table 3.5, we postulated a multiple logistic regression model with backward selection process to estimate π_{im}. The logistic regression model can be written as follows: $\hat{\pi}_{im} = G_{im}(\hat{\boldsymbol{\gamma}}) = [1 + \exp(\mathbf{O}_{im}^T\hat{\boldsymbol{\gamma}})]^{-1}$, where $\mathbf{O}_{im}^T = \{I(t_{im} = t_{i2}), \ldots, I(t_{im} = t_{i13}), \text{edu}_i, \text{HRSD}_{i(m-1)}, \text{age}_i, \text{HRSD}_{i(m-1)} \times \text{age}_i\}$.

Table 3.6 shows the results of analyzing data from the REVAMP study. We performed the complete-case (CC) and the available-case (AC) with IPWGEE analyses. The standard error of the estimates from the AC with IPWGEE were slightly smaller than those from the CC. Controlling for the employment status, for a patient of age 43 years, the treatment regime SSB will have the highest reduction of 0.759 per week in the HRSD scores from baseline, followed by the treatment regime SSC (0.693/week) and then by the treatment regime SSM (0.559/week) in the CC analysis. On the contrary, the treatment regime SSB had the highest reduction of 0.780 per week in the HRSD scores from baseline, followed by the treatment regime SSM (0.729/week) and then by the treatment regime SSC (0.710/week) in the AC with IPWGEE analysis. In both analyses, the effect of each treatment regime was statistically significant (p<0.001). To compare the effects across treatment regimes SSC, SSB, and SSM, we used the Wald test discussed in Section 2.5. In Table 3.7, Wald Chi-square tests comparing the effects of treatment regimes ($H_0 : \beta_{1,SSC} + \beta_{4,SSC} \times 43 = \beta_{1,SSB} + \beta_{4,SSB} \times 43 = \beta_{1,SSM} + \beta_{4,SSM} \times 43$) with 2 degrees of freedom resulted in a p-values of 0.500 in the CC analysis and 0.363 in the AC with IPWGEE analysis. Both methods indicated that there was no evidence that the effects of these three treatment regimes were significantly different from each other.

3.6 DISCUSSION

Missing data is a common phenomenon in longitudinal studies. Missing data can be broadly attributed to two sources: (1) study design or randomization, and (2) drop-out from the study. In a two-stage longitudinal study, such as the one presented here, in the second stage of the study, patients randomized to one treatment can not receive other competing treatments. Additionally, if the drop-out occurs prior to the second stage, a patient's response status may not be observed. To account for these two sources of missing data, we have used the inverse-probability-weighted generalized estimating equations (IPWGEE) method (Robins et al., 1995) to take into account the missing data due to randomization and drop-out. The weights are formed by inversely weighting the probability of randomization to the treatment dictated by the regime and the probability of observing outcome data at each visit. In this Chapter, we have constructed time-dependent weights to take into account the partial information from patients who received other competing treatments and patients who dropped out from the study. The probability of observing outcome data at each visit is estimated through a logistic regression model. We show that the IPWGEE estimators are consistent and asymptotically normal. We have demonstrated our methods using a dataset from a depression study.

3.7 FIGURES AND TABLES

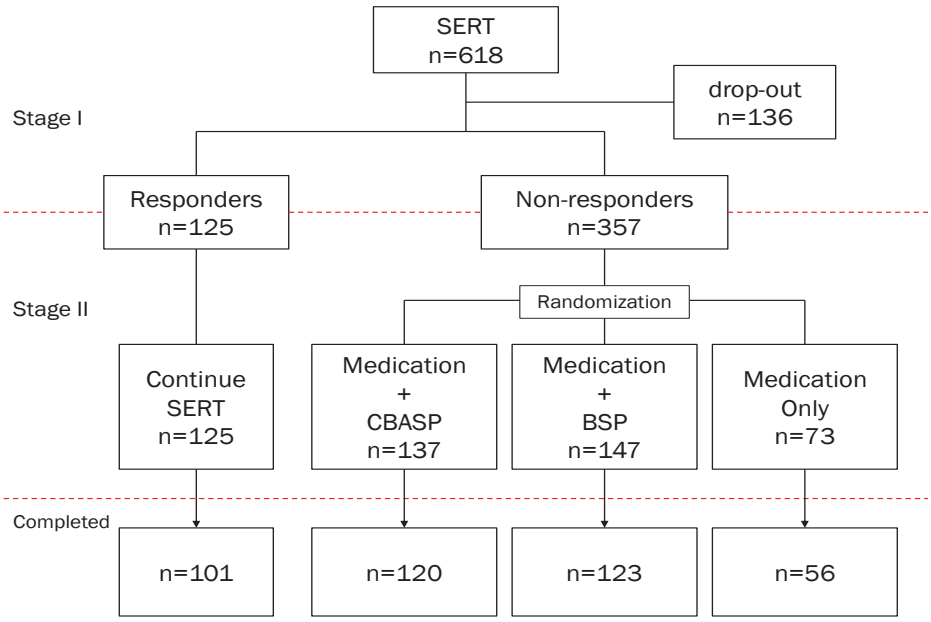
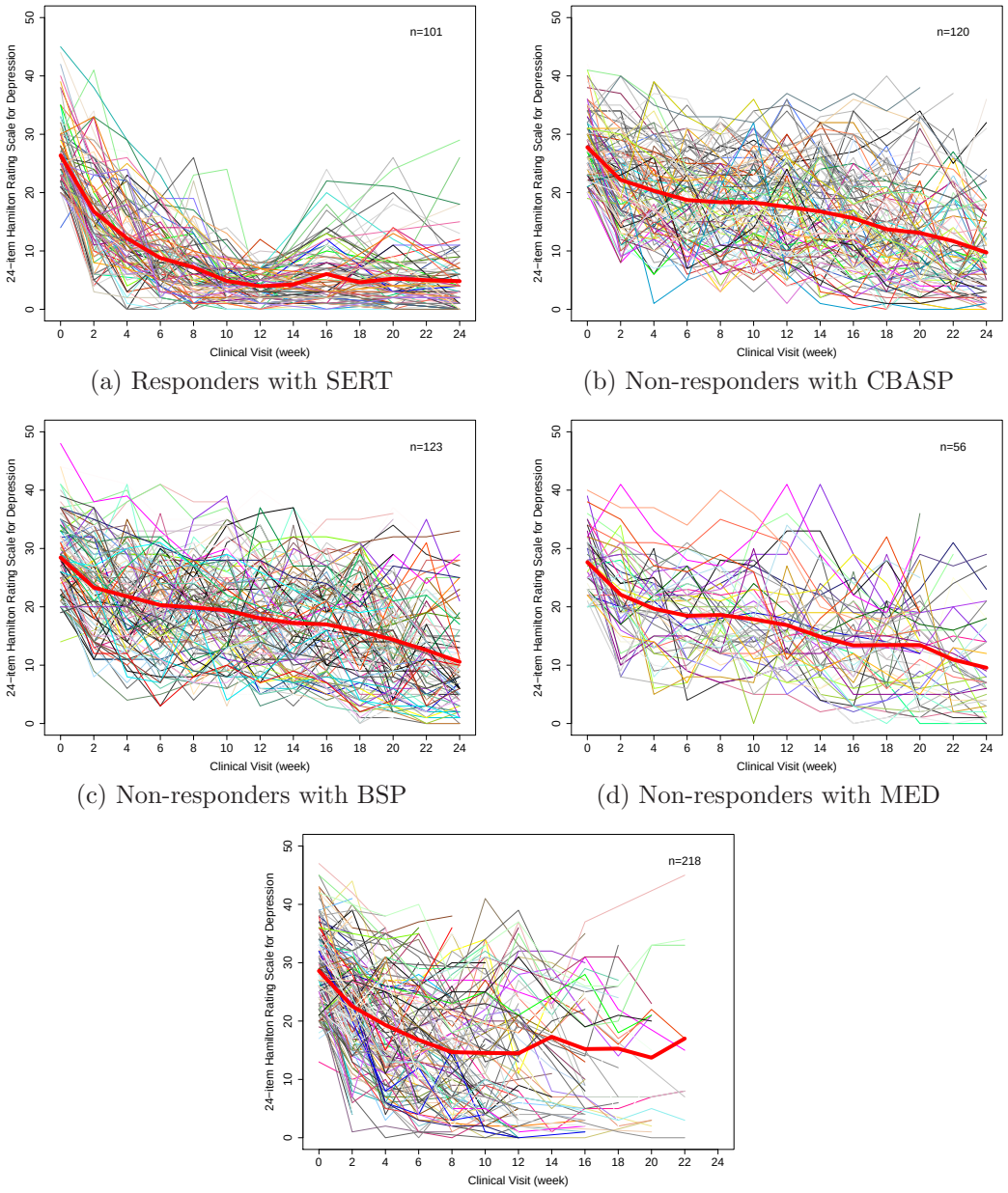


Figure 3.1: **Patient flow in the REVAMP study.**



(a) Responders with SERT

(b) Non-responders with CBASP

(c) Non-responders with BSP

(d) Non-responders with MED

(e) Drop-outs

Figure 3.2: **HRSD scores of patients in the REVAMP study.** Thick lines indicate the mean of HRSD scores at each visit. (CBASP: Cognitive Behavior Analysis System of Pychotherapy; BSP: Brief Supportive Psychotherapy; MED: Medication alone)

Table 3.1: **Simulation results of $\hat{\beta}_{1,1kl}$ based on 2000 Monte Carlo samples of sizes 250, 500, and 900 with 50% response rate.** EST is MC mean of estimates, SE is MC mean of estimated standard errors, MCSE is standard error of MC estimates, and CP is empirical coverage probability.

Sample Size	Regime	True Value	EST (SE)	MCSE	CP%	EST (SE)	MCSE	CP%
			30% drop-out rate			50% drop-out rate		
250	$A_1B_1B_1'$	-1.13	-1.13 (0.050)	0.052	94.0	-1.12 (0.074)	0.083	91.3
	$A_1B_1B_2'$	-1.53	-1.52 (0.085)	0.086	94.9	-1.51 (0.133)	0.145	91.0
	$A_1B_2B_1'$	-1.30	-1.29 (0.048)	0.049	94.7	-1.29 (0.070)	0.078	91.8
	$A_1B_2B_2'$	-1.69	-1.68 (0.070)	0.070	94.1	-1.68 (0.109)	0.117	92.2
500	$A_1B_1B_1'$	-1.13	-1.13 (0.036)	0.036	94.5	-1.13 (0.054)	0.058	92.4
	$A_1B_1B_2'$	-1.53	-1.52 (0.061)	0.065	93.2	-1.51 (0.097)	0.104	91.3
	$A_1B_2B_1'$	-1.30	-1.29 (0.034)	0.035	93.5	-1.29 (0.050)	0.055	92.4
	$A_1B_2B_2'$	-1.69	-1.69 (0.050)	0.053	94.3	-1.68 (0.079)	0.085	91.3
900	$A_1B_1B_1'$	-1.13	-1.13 (0.027)	0.027	94.3	-1.13 (0.040)	0.041	94.3
	$A_1B_1B_2'$	-1.53	-1.52 (0.045)	0.047	93.5	-1.51 (0.074)	0.077	92.1
	$A_1B_2B_1'$	-1.30	-1.29 (0.026)	0.027	93.5	-1.29 (0.038)	0.039	94.0
	$A_1B_2B_2'$	-1.69	-1.69 (0.037)	0.038	93.5	-1.68 (0.059)	0.062	91.9

Table 3.2: **Simulation results of $\hat{\beta}_{1,1kl}$ based on 2000 Monte Carlo samples of sizes 250, 500, and 900 with 30% response rate.** EST is MC mean of estimates, SE is MC mean of estimated standard errors, MCSE is standard error of MC estimates, and CP is empirical coverage probability.

Sample Size	Regime	True Value	EST (SE)	MCSE	CP%	EST (SE)	MCSE	CP%
			30% drop-out rate			50% drop-out rate		
250	$A_1B_1B_1'$	-1.21	-1.20 (0.051)	0.052	93.8	-1.20 (0.075)	0.076	95.3
	$A_1B_1B_2'$	-1.76	-1.74 (0.082)	0.084	93.7	-1.72 (0.132)	0.145	91.7
	$A_1B_2B_1'$	-1.31	-1.29 (0.049)	0.050	93.1	-1.30 (0.071)	0.075	94.5
	$A_1B_2B_2'$	-1.86	-1.84 (0.069)	0.072	92.5	-1.83 (0.109)	0.120	89.5
500	$A_1B_1B_1'$	-1.21	-1.20 (0.036)	0.037	93.0	-1.19 (0.054)	0.060	91.2
	$A_1B_1B_2'$	-1.76	-1.74 (0.059)	0.055	94.6	-1.73 (0.095)	0.097	93.1
	$A_1B_2B_1'$	-1.31	-1.29 (0.035)	0.037	91.1	-1.29 (0.051)	0.055	92.1
	$A_1B_2B_2'$	-1.86	-1.84 (0.049)	0.046	91.7	-1.84 (0.078)	0.081	92.5
900	$A_1B_1B_1'$	-1.21	-1.20 (0.027)	0.028	92.9	-1.19 (0.041)	0.041	94.3
	$A_1B_1B_2'$	-1.76	-1.74 (0.044)	0.043	93.3	-1.73 (0.072)	0.076	91.7
	$A_1B_2B_1'$	-1.31	-1.29 (0.026)	0.027	91.9	-1.29 (0.039)	0.040	92.0
	$A_1B_2B_2'$	-1.86	-1.84 (0.037)	0.036	91.7	-1.83 (0.059)	0.062	91.9

Table 3.3: **Comparisons of subject-specific and time-dependent weights for estimating $\beta_{1,1kl}$ based on 2000 Monte Carlo samples of 50% response rate.** EST is MC mean of estimates and MCSE is standard error of MC estimates.

			Subject-Specific Weight				Time-Dependent Weight			
			30% Drop-out		50% Drop-out		30% Drop-out		50% Drop-out	
Size	Regime	Truth	EST	MCSE	EST	MCSE	EST	MCSE	EST	MCSE
250	$A_1B_1B_1'$	-1.13	-1.13	0.070	-1.13	0.136	-1.13	0.052	-1.12	0.083
	$A_1B_1B_2'$	-1.53	-1.53	0.099	-1.53	0.182	-1.52	0.086	-1.51	0.145
	$A_1B_2B_1'$	-1.30	-1.30	0.064	-1.29	0.128	-1.29	0.049	-1.29	0.078
	$A_1B_2B_2'$	-1.69	-1.70	0.083	-1.69	0.152	-1.68	0.070	-1.68	0.117
500	$A_1B_1B_1'$	-1.13	-1.13	0.050	-1.13	0.100	-1.13	0.036	-1.13	0.058
	$A_1B_1B_2'$	-1.53	-1.53	0.072	-1.53	0.131	-1.52	0.065	-1.51	0.104
	$A_1B_2B_1'$	-1.30	-1.30	0.047	-1.30	0.093	-1.29	0.035	-1.29	0.055
	$A_1B_2B_2'$	-1.69	-1.70	0.059	-1.69	0.111	-1.69	0.053	-1.68	0.085
900	$A_1B_1B_1'$	-1.13	-1.13	0.037	-1.14	0.067	-1.13	0.027	-1.13	0.041
	$A_1B_1B_2'$	-1.53	-1.53	0.053	-1.53	0.100	-1.52	0.047	-1.51	0.077
	$A_1B_2B_1'$	-1.30	-1.30	0.034	-1.30	0.059	-1.29	0.027	-1.29	0.039
	$A_1B_2B_2'$	-1.69	-1.70	0.044	-1.70	0.078	-1.69	0.038	-1.68	0.062

Table 3.4: **Comparisons of subject-specific and time-dependent weights for estimating $\beta_{1,1kl}$ based on 2000 Monte Carlo samples of 30% response rate.** EST is MC mean of estimates and MCSE is standard error of MC estimates.

			Subject-Specific Weight				Time-Dependent Weight			
			30% Drop-out		50% Drop-out		30% Drop-out		50% Drop-out	
Size	Regime	Truth	EST	MCSE	EST	MCSE	EST	MCSE	EST	MCSE
250	$A_1B_1B_1'$	-1.21	-1.20	0.071	-1.21	0.132	-1.20	0.052	-1.20	0.076
	$A_1B_1B_2'$	-1.76	-1.76	0.094	-1.76	0.174	-1.74	0.084	-1.72	0.145
	$A_1B_2B_1'$	-1.31	-1.30	0.067	-1.31	0.135	-1.29	0.050	-1.30	0.075
	$A_1B_2B_2'$	-1.86	-1.86	0.082	-1.86	0.152	-1.84	0.072	-1.83	0.120
500	$A_1B_1B_1'$	-1.21	-1.21	0.045	-1.20	0.099	-1.20	0.037	-1.19	0.060
	$A_1B_1B_2'$	-1.76	-1.76	0.058	-1.76	0.124	-1.74	0.055	-1.73	0.097
	$A_1B_2B_1'$	-1.31	-1.30	0.046	-1.30	0.091	-1.29	0.037	-1.29	0.055
	$A_1B_2B_2'$	-1.86	-1.86	0.052	-1.86	0.104	-1.84	0.046	-1.84	0.081
900	$A_1B_1B_1'$	-1.21	-1.20	0.037	-1.20	0.069	-1.20	0.028	-1.19	0.041
	$A_1B_1B_2'$	-1.76	-1.76	0.050	-1.76	0.097	-1.74	0.043	-1.73	0.076
	$A_1B_2B_1'$	-1.31	-1.30	0.035	-1.30	0.064	-1.29	0.027	-1.29	0.040
	$A_1B_2B_2'$	-1.86	-1.86	0.042	-1.86	0.080	-1.84	0.036	-1.83	0.062

Table 3.5: **Drop-out rates by baseline characteristics.**

Characteristics	Complete (n=400) n (%)	Drop-out (n=218) n (%)	p-value
Age (years)			
Age $\leq$ 46	185 (53.9)	158 (46.1)	<0.01
Age > 46	214 (78.1)	60 (21.9)	
Missing, No.	1	0	
Sex			
Male	185 (66.1)	95 (33.9)	0.52
Female	215 (63.6)	123 (36.4)	
Caucasian Race			
Yes	344 (66.9)	170 (33.1)	0.01
No	56 (53.9)	48 (46.2)	
Hispanic ethnicity			
Yes	33 (58.9)	23 (41.1)	0.34
No	362 (65.3)	192 (34.7)	
Missing, No.	5	3	
Employment			
Employed	255 (66.9)	126 (33.1)	0.12
Unemployed	141 (60.8)	91 (39.2)	
Missing, No.	4	1	
Education			
$\leq$ High school	139 (53.9)	119 (46.1)	<0.01
> High school	241 (71.5)	96 (28.5)	
Missing, No.	20	3	
Marital status			
Married	158 (68.7)	72 (31.3)	0.10
Not married	238 (62.1)	145 (37.9)	
Missing, No.	4	1	

Table 3.6: **Results of estimating the effects of treatment regimes in the REVAMP study using complete-case and IPWGEE analyses.**

		Complete-Case			IPWGEE		
Treatment Regime	Effect	Estimated Parameter	Standard Error	p-value	Estimated Parameter	Standard Error	p-value
SSC [1]	Intercept	27.192	1.676	<0.001	26.042	1.332	<0.001
	Time	-0.865	0.094	<0.001	-0.882	0.093	<0.001
	Age	0.015	0.030	0.627	-0.011	0.025	0.647
	Employment	-1.783	0.895	0.046	-2.030	0.845	0.016
	Time×Age	0.004	0.002	0.069	0.004	0.002	0.026
SSB [2]	Intercept	27.185	1.517	<0.001	26.369	1.127	<0.001
	Time	-0.931	0.079	<0.001	-0.955	0.072	<0.001
	Age	-0.010	0.028	0.705	0.004	0.021	0.853
	Employment	-2.184	0.772	0.005	-2.049	0.670	0.002
	Time×Age	0.004	0.002	0.013	0.005	0.002	0.002
SSM [3]	Intercept	27.324	2.057	<0.001	25.824	1.707	<0.001
	Time	-0.645	0.111	<0.001	-0.730	0.107	<0.001
	Age	-0.007	0.034	0.830	0.019	0.029	0.517
	Employment	-2.706	1.087	0.013	-2.044	1.087	0.060
	Time×Age	0.002	0.002	0.425	0.00003	0.002	0.990

[1] Treat with SERT, continue SERT if respond, otherwise add CBASP to SERT

[2] Treat with SERT, continue SERT if respond, otherwise add BSP to SERT

[3] Treat with SERT, continue SERT if respond, otherwise add MED to SERT

Table 3.7: **Comparisons of estimated time effects of treatment regimes for a patient of 43 years controlling for employment status in the REVAMP study.**

Method	Regime	Estimated Effect	χ^2_2	p-value
Complete-Case	SSC [1]	-0.865 + 0.004×43	1.384	0.500
	SSB [2]	-0.931 + 0.004×43		
	SSM [3]	-0.645 + 0.002×43		
IPWGEE	SSC [1]	-0.882 + 0.004×43	2.028	0.363
	SSB [2]	-0.955 + 0.005×43		
	SSM [3]	-0.730 + 0.00003×43		

[1] Treat with SERT, continue SERT if respond, otherwise add CBASP to SERT

[2] Treat with SERT, continue SERT if respond, otherwise add BSP to SERT

[3] Treat with SERT, continue SERT if respond, otherwise add MED to SERT

4.0 CONCLUSION

4.1 SUMMARY

In this dissertation, we have demonstrated the estimation of the effects of treatment regimes in two-stage longitudinal studies in the presence of missing data. To estimate the effect of a regime from a two-stage longitudinal study with missing data, we have adapted the inverse-probability-weighted generalized estimating equations (IPWGEE) method to correct the bias that is caused by missing data.

In Chapter 2, we constructed the weights based on the probability of receiving treatment and the probability of having complete data at patient level. The probability of receiving treatment was known by study design and the probability of having complete data was estimated from the sample through a logit model. We showed that under certain assumptions, the IPWGEE estimators are consistent and asymptotically normal. We also showed how to compare the treatment regimes via the Wald test, which required computation of covariance between two estimated regime effects.

In Chapter 3, we have extended the results one step further from Chapter 2. Previously, patients with incomplete data were ignored. In Chapter 3, we incorporated the information on the time to response and the time to drop-out. We constructed the weights at visit level. This allowed us to utilize more observations from patients who dropped out from the study or patients who were not consistent with the treatment regime to estimate the parameters of interest with larger efficiency.

4.2 FUTURE WORK

Our proposed IPWGEE estimators in two-stage longitudinal studies provided valid estimation when the missing data mechanism is missing at random. The efficiency of the proposed IPWGEE estimators can be improved by incorporating partially observed information from patients who have missing data due to randomization and drop-out.

The IPWGEE method used in Chapter 2 and 3 require the assumption that both marginal mean model and missing data model to be correctly specified. However, the combination of the two model specification requirements will increase the chance of at least one misspecification and lead to bias. Doubly robust estimating equations [Scharfstein et al., 1999, van der Laan and Robins, 2003, Bang and Robins, 2005] can provide unbiased estimators as long as either one of the preceding models is correctly specified. Developing doubly robust weighted estimating equations in two-stage longitudinal studies will be the basis of our future research.

4.3 PUBLIC HEALTH SIGNIFICANCE

Mental illness is becoming a major public health challenge. Strategies of multiple treatments have been introduced by many investigators to serve as an alternative to single strategy in treating patients with chronic depressive disorders. As the complexity of study design increases, developing sophisticated statistical method is necessary in order to provide valid inference. This dissertation demonstrates the importance of statistical aspects to estimate the effects of depression treatment regimes from two-stage longitudinal studies.

BIBLIOGRAPHY

Heejung Bang and James M. Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61:962–972, 2005.

Oliver Bembom and Mark J. van der Laan. Analyzing sequentially randomized trials based on causal effect models for realistic individualized treatment rules. *Statistics in Medicine*, 27:3689–3716, 2008.

Lisa M. Bodnar, Marie Davidian, Anna Maria Siega-Riz, and Anastasios A. Tsiatis. Marginal structural models for analyzing causal effects of time-dependent treatments: An application in perintal epidemiology. *American Journal of Epidemiology*, 159(10):926–934, 2004.

Pim Cuijpers, Annemieke van Straten, Lisanne Warmerdam, and Gerhard Andersson. Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: A meta-analysis. *Depression and Anxiety*, 26:279–288, 2009.

Pim Cuijpers, Annemieke van Straten, Josien Schuurmans, Patricia van Oppen, Steven D. Hollon, and Gerhard Andersson. Psychotherapy for chronic major depression and dysthymia: A meta-analysis. *Clinical Psychology Review*, 30:51–62, 2010.

Peter J. Diggle, Patrick Heagerty, Kung-Yee Liang, and Scott L. Zeger. *Analysis of longitudinal data*. Oxford University Press, New York, 2002.

Max Hamilton. A rating scale for depression. *Journal of Neurology, Neurosugery, & psychiatry*, 23:45–62, 1960.

Miguel Á. Hernán, Babette Brumback, and James M. Robins. Marginal structural models to estimate the causal effect of zidovudine on the survival of hiv-positive men. *Epidemiology*, 11(5):561–570, 2000.

Miguel Angel Hernán. A definition of causal effect for epidemiological research. *Journal Epidemiological Community Health*, 58:265–271, 2004.

Paul W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.

D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

Martin B. Keller, James P. McCullough, Daniel N. Klein, Bruce Arnow, David L. Dunner, Alan J. Gelenberg, John C. Markowitz, Charles B. Nemeroff, James M. Russell, Michael E. Thase, Madhukar H. Trivedi, and John Zajecka. A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *The New England Journal of Medicine*, 342(20):1462–1464, 2000.

Daniel N. Klein, Bruce A. Arnow, Jennifer L. Barkin, Frank Dowling, James H. Kocsis, Andrew C. Leon, Rachel Manber, Barbara O. Rothbaum, Madhukar H. Trivedi, and Stephen R. Wisniewski. Early adversity in chronic depression: Clinical correlates and response to pharmacotherapy. *Depression and Anxiety*, 26:701–710, 2009.

Hyejin Ko, Joseph W. Hogan, and Kenneth H. Mayer. Estimating causal treatment effects from longitudinal hiv natural history studies using marginal structural models. *Biometrics*, 59:152–162, 2003.

Jin Hui Ko. *Statistical issues in the design and analysis of sequentially randomized trials*. PhD thesis, University of Pittsburgh, 2010.

James H. Kocsis, Alan J. Gelenberg, Barbara O. Rothbaum, Daniel N. Klein, Madhukar H. Trivedi, Rachel Manber, Martin B. Keller, Robert howland, and Michael E. Thase. Chronic forms of major depression are still undertreated in the 21st century: Systematic assessment of 801 patients presenting for treatment. *Journal of Affective Disorders*, 110:55–61, 2008.

James H. Kocsis, Alan J. Gelenberg, Barbara O. Rothbaum, Daniel N. Klein, Madhukar H. Trivedi, Rachel Manber, Martin B. Keller, Andrew C. Leon, Stephen R. Wisniewski, Bruce A. Arnow, John C. Markowitz, Michael E. Thase, and REVAMP investigators. Cognitive behavioral analysis system of psychotherapy and brief supportive psychotherapy for augmentation of antidepressant nonresponse in chronic depression: the revamp trial. *Archives of general psychiatry*, 66(11):1178–88, 2009.

Philip W. Lavori and Ree Dawson. A design for testing clinical strategies: biased adaptive within-subject randomization. *Journal of the Royal Statistical Society: Series A*, 163(1): 29–38, 2000.

Philip W. Lavori, A. John Rush, Stephen R. Wisniewski, Jonathan Alpert, Maurizio Fava, David J. Kupfer, Andrew Nierenberg, Frederic M. Quitkin, Harold A. Sackeim, Michael E. Thase, and Madhukar Trivedi. Strengthening clinical effectiveness trials: Equipoise-stratified randomization. *Biological Psychiatry*, 50(10):792–801, 2001.

Kung-Yee Liang and Scott L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

Roderick J. A. Little and Donald B. Rubin. *Statistical analysis with missing data*. Wiley, New York, 2002.

Jared K. Lunceford, Marie Davidian, and Anastasios A. Tsiatis. Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics*, 58:48–57, 2002.

Susan A. Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 65(2):331–366, 2003.

Susan A. Murphy. An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*, 24:1455–1481, 2005.

Susan A. Murphy, Mark J. van der Laan, James M. Robins, and Conduct Problems Prevention Research Group. Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*, 96(456):1410–1423, 2001.

Andrew A. Nierenberg, Timothy J. Petersen, and Jonathan E. Alpert. Prevention of relaspse and recurrence in depression: The role of long-term pharmacotherapy and psychotherapy. *Journal of Clinical Psychiatry*, 64(suppl 15):13–17, 2003.

James M. Robins. A new approach to causal inference in mortality studies with a sustained exposure period – application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7:1393–1512, 1986.

James M. Robins. Addendum to "a new approach to causal inference in mortality studies with a sustained exposure period – application to control of the healthy worker survivor effect.". *Computers & Mathematics with Applications.*, 14(9–12):923–=945, 1987.

James M. Robins. Causal inference from complex longitudinal data. In M. Berkane, editor, *Latent Variable Modeling and Applications to Causality*, pages 69–117. Springer, New York, 1997.

James M. Robins and Andrea Rotnitsky. Recovery of information and adjustment for dependent censoring using surrogate markers. In Nicholas P. Jewell, Klaus Dietz, and Vernon T. Farewell, editors, *AIDS Epidemiology – Methodological Issues.*, pages 297–331. Birkhäuser, Boston, 1992.

James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429):106–121, 1995.

James M. Robins, Miguel Angel Hernán, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

Andrea Rotnitzky. Inverse probability weighted methods. In Garrett Fitzmaurice, Marie Davidian, Geert Verbeke, and Geert Molenberghs, editors, *Longitudinal data analysis.*, pages 453–476. CRC/Chapman & Hall, Boca Raton, FL, 2009.

Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

Donald B. Rubin. Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 6(1):34–58, 1978.

Daniel O. Scharfstein, Andrea Rotnitzky, and James M. Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120, 1999.

Jerzy Splawa-Neyman. On the application of probability theory to agricultural experiments (translated and edited by d.m. dabrowska and t.p. speed). *Statistical Science*, 5(4):465–480, 1990.

Leonard A. Stefanski and Dennis D. Boos. The calculus of m-estimation. *The American Statistician*, 56(1):29–38, 2002.

Thomas R. TenHave, James Coyne, Mark Salzer, and Ira Katz. Research to improve the quality of care for depression: alternatives to the simple randomized clinical trial. *General Hospital Psychiatry*, 25:115–123, 2003.

Peter F. Thall, Randall E. Millikan, and Hsi-Guang Sung. Evaluating multiple treatment courses in clinical trials. *Statistics in Medicine*, 19:1011–1028, 2000.

Peter F. Thall, Hsi-Guang Sung, and Elihu H. Estey. Selecting therapeutic strategies based on efficacy and death in multicourse clinical trials. *Journal of the American Statistical Association*, 97(457):29–39, 2002.

Michael E. Thase, Edward S. Friedman, Melanie M. Biggs, Stephen R. Wisniewski, Madhukar H. Trivedi, James F. Luther, Maurizio Fava, Andrew A. Nierenberg, Patrick J. McGrath, Diane Warden, George Niederehe, Steven D. Hollon, and A. John Rush. Cognitive therapy versus medication in augmentation and switch strategies as second-step treatments: A star*d report. *The American Journal of Psychiatry*, 164:739–752, 2007.

Madhukar H. Trivedi, James H. Kocsis, Michael E. Thase, David W. Horris, Stephen R. Wisniewski, Andrew C. Leon, Alan J. Gelenberg, Daniel N. Klein, George Niederehe, Alan F. Schatzberg, Philip T. Ninan, and Martin B. Keller. Revamp - research evaluating the value of augmenting medication with psychotherapy: rationale and design. *Psychopharmacology Bulletin*, 41(4):5–33, 2008.

Mark J. van der Laan and James M. Robins. *Unified methods for censored longitudinal data and causality*. Springer, New York, 2003.

Taylor J. VanderWeele. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*, 20(1):18–26, 2009.

Scott L. Zeger and Kung-Yee Liang. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42(1):121–130, 1986.